# Essays on Education and Family Behaviour

**Ziyi Huang**

A thesis submitted for the degree of

*Doctor of Philosophy*

December 2025

University of Essex

# Author's Declaration

I confirm that this thesis has received ethical approval from the University of Essex for the use of non-public data. The approval reference numbers are *ETH2223-2269* and *ETH2223-2338*.

I confirm that this thesis contains both independent work undertaken by myself and joint work collaborated with other researchers.

The co-authored work includes the following:

- **Chapter 2: Effect of School Meal Policy on Parents' Outcomes: Evidence from China**

  My contribution to this work was 90%, which comprised the framing and conceptualisation of the study, implementation of the econometric model, literature review, data analysis, and writing of the chapter.

  Qingxu Yang contributed 10%, including data collection and providing information on variable definitions and dataset structure.

- **Chapter 3: Absences from school and educational outcomes. Religious observance among ethnic minority students**

  My contribution to this work was 60%, which included implementing the econometric model, reviewing literature, analysing the results, and drafting the chapter.

  Birgitta Rabe contributed 20%, including framing and conceptualisation, model design, and writing of the introduction.

  Angus Holford contributed 20%, including framing and conceptualisation, model design, and writing of the introduction.

All other chapters and analyses presented in this thesis are entirely my own work.

**Signed:** Ziyi Huang

**Date:** 11 October 2025

# Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr Angus Holford, for his patience, guidance, and generous support throughout my research. His supervision has not only greatly improved the quality of this thesis but has also strengthened my research abilities. He has provided comprehensive assistance, offering detailed feedback on my chapters, insightful advice on academic writing, and continuous encouragement that helped me integrate into the research community at the University. I am deeply grateful for all his help, which has been the most important part of my PhD experience.

I would like to express my sincerest thanks to my co-supervisor, Professor Birgitta Rabe, for her kindness, guidance, and professional support during my studies. She offered different and valuable perspectives that helped me understand my topics better. She also helped me get familiar with the local culture and research community.

I would also express my thanks to my panel chairs, Professor Emilia Del Bono and Dr Laura Fumagalli, as well as the academic team in ISER. I am very grateful for their advice and continuous support on my research. I would also like to thank ISER for providing an excellent research environment and an inspiring academic community, which made my PhD experience both enjoyable and fulfilling.

# Abstract

This thesis explores children's education and related family behaviours, focusing on the interactions among children, parents, and schools. It uses large-scale survey and administrative data from China and England to study these relationships from three perspectives.

Chapter 1 examines effect of parents' inputs, including family and private tutoring, on students' educational outcomes. Using data from the CEPS, the results show that tutoring has limit positive effects on school exam performance, and negative effects on cognitive ability. Further analysis suggests that private tutoring strengthens test-taking ability rather than cognitive skills, indicating that exam-oriented tutoring brings short-term gains but may harm cognitive ability development.

Chapter 2 investigates how school meal policies influence parents' time allocation in leisure and labour market. Using DID imputation method with CGSS and CHFS data, this chapter finds that providing school meals reallocate parents' time and increase their time spending in labour market. Our results in heterogeneity suggest that that providing school meals narrows socioeconomic gaps in parents' welfare.

Chapter 3 analyses how religious festivals that coincide with school days affect students' attendance and academic performance. Using data from the National Pupil Database, the study applies an instrumental variable strategy based on the exogenous timing of religious festivals determined by the lunar calendar. The findings show that religious festivals falling on school days increases students' absences, and that these additional absences negatively impact students' educational outcomes. Our research suggests that adjusting the school term dates can lower pupils' absence rates and improve educational attainment in districts with a large number of students with specific religions.

# Introduction

Although there is debate about what children should be taught, the importance of education itself is beyond question for most people. For children, education shapes their character, develops their abilities, and lays the foundation for their future participation in society. For households, children are the future of the family, and education is both a parental duty and a long-term investment. For society, education is more than the labour supply for the future, and it also determines the shape of future society.

There are three main participants in the educational process. Besides the children who receive education, families and schools are often the primary investors. This thesis examines the relationships among students, parents, and schools. It investigates the effect of parental investment on students, the impact of school policies on parents, and the influence of these policies on children.

The analysis in this thesis shows that parents and schools have both direct and indirect effects on students. Parents can invest in education by spending time on family tutoring or by paying for private tutoring. Some parental characteristics, such as their religion, can also affect children's academic performance. When the religion of parents differs from the major religion in the school, their children with the same belief may be absent for religious reasons, which can lower their grades. From the school's perspective, certain policies, such as the school calendar and absence rules, directly affect attendance and performance. Other policies, such as school meal programs, can improve students' nutritional status and physical health while also reducing parents' time or financial burdens. This enables parents to invest more resources in their children's education.

Each chapter of the thesis explores these links in detail. The first chapter studies the link between parents and students and is titled *"Parental Investment and Cognitive Outcomes in Adolescence: Evidence from China."* This chapter uses longitudinal data from the China Education Panel Survey (CEPS) and multiple empirical strategies to examine how family tutoring and private tutoring affect students' examination ranks and cognitive test ranks. The results show that family tutoring has no positive effects on students' examination or cognitive performance. Private tutoring improves school test rankings in mathematics but shows no benefits for cognitive outcomes. Further analysis indicates that private tutoring affects examination ranks by enhancing students' test-taking ability rather than developing cognitive skills. The results in this chapter suggest that tutoring institutions specialize in exam-oriented strategies that yield short-term advantages in high-stakes tests but may crowd out the development of cognitive abilities.

The second chapter, titled *"Effect of School Meal Policy on Parents' Outcomes: Evidence from China,"* studies the link between schools and parents. This chapter uses data from the Chinese General Social Survey (CGSS) and the China Household Finance Survey (CHFS). By implementing a Difference-in-Differences (DiD) estimation method, this chapter tests the effects of the roll-out of school meals and subsidized meals on parents' leisure time allocation and labour market participation in China. The findings provide evidence that school meal policies can have spillover effects beyond their direct impact on children, especially releasing parents' time. After the implementation of the school meal policy, urban *hukou* fathers spent more time listening to music, while rural *hukou* parents increased the time spent watching TV. In the labour market, both urban and rural *hukou* parents increased their weekly working hours. In further heterogeneity analysis, this chapter suggests that providing school meals narrows socioeconomic gaps in parents' welfare

The third chapter, titled *"Absences from school and educational outcomes. Religious observance among ethnic minority students"*, studies the link between schools and students. This chapter uses data from the National Pupil Database and implements an instrumental variable (IV) approach, where the IV is religious festivals falling on school days. The results show that religious festivals falling on school days increase students' absences, and these additional absences negatively affect students' educational outcomes. Therefore, this chapter shows that family religious background influences students' academic outcomes indirectly.

By revealing these direct and indirect effects, the thesis provides new empirical evidence and policy insights for understanding the interactions among different participants in education. First, it shows that the effects of educational policies extend beyond students and can indirectly affect parents' time allocation. Policy evaluation should therefore go beyond a single dimension. Second, the findings highlight both the efficiency and the limitations of parental investment within an exam-oriented education system. Policymakers should consider long-term skill development when designing the allocation of educational resources and regulating private tutoring. Finally, this thesis demonstrates how school institutional design can influence family welfare and educational equity through policies such as meal provision and holiday arrangements. This offers empirical support for creating a more inclusive and equitable educational system.

# Contents

**1 Parental investment and cognitive outcomes in adolescence: evidence from China 11**

1.1 Introduction . . . 12
1.2 Literature Review . . . 14
1.2.1 International Evidence . . . 14
1.2.2 Evidence from China . . . 15
1.2.3 Distinguishing School Test Scores from Cognitive Test Scores . . . 16
1.2.4 Contribution . . . 17
1.3 Data . . . 17
1.3.1 China Education Panel Survey . . . 17
1.3.2 Outcome Variables in CEPS . . . 18
1.3.3 Test-taking ability . . . 19
1.3.4 Private Tutoring (PT) . . . 21
1.3.5 Family Tutoring (FT) . . . 23
1.3.6 Control Variables . . . 24
1.3.7 Descriptive statistics . . . 24
1.4 Methods . . . 30
1.4.1 Contemporaneous Model . . . 31
1.4.2 Two-Way Fixed Effects Model (TWFE) . . . 32
1.4.3 Value-Added Model (VA) . . . 33
1.4.4 Cumulative model (CU) . . . 34
1.4.5 Cumulative Value-Added Model (CVA) . . . 34
1.5 Estimation Results . . . 35

1.5.1 School Test Rank . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

1.5.2 Cognitive ability VS Test-taking ability . . . . . . . . . . . . . . . . . 37

1.6 Robustness check . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

1.7 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50

**2 Effect of school meal policy on parents' outcomes: evidence from China 51**

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

2.2 Literature . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

2.2.1 Types of school meal treatment effects . . . . . . . . . . . . . . . . . 55

2.2.2 Effects on students and other family members . . . . . . . . . . . . . 57

2.3 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58

2.3.1 Schooling in China . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58

2.3.2 Early practice of school meal in China . . . . . . . . . . . . . . . . . 59

2.3.3 The Student Nutrition Improvement Program (SNIP) . . . . . . . . . . 59

2.4 Definition of the treatment . . . . . . . . . . . . . . . . . . . . . . . . . . 61

2.4.1 Policies as treatments and school canteen control . . . . . . . . . . . . 61

2.4.2 Exposure to SNIP by hukou type . . . . . . . . . . . . . . . . . . . . . 62

2.4.3 Details of treatment . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63

2.5 Framework . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

2.5.1 Impact on household's lunch decision . . . . . . . . . . . . . . . . . . 64

2.5.2 Impact of lunch decision change on household's constraints . . . . . . 65

2.6 Data and sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

2.6.1 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

2.6.2 Observations in estimations . . . . . . . . . . . . . . . . . . . . . . . 67

2.6.3 Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69

2.6.4 Descriptive Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

2.7 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

2.7.1 Regression and estimator . . . . . . . . . . . . . . . . . . . . . . . . . 74

2.7.2 Identification Assumptions . . . . . . . . . . . . . . . . . . . . . . . . 76

2.7.3 Parallel Trends . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
2.7.4 Potential issues in identification . . . . . . . . . . . . . . . . . . . . 77
2.8 Average impacts and heterogeneity by duration . . . . . . . . . . . . . . . . 79
2.8.1 Reallocation of time in leisure activities . . . . . . . . . . . . . . . . 80
2.8.2 Reallocation of time in labour market . . . . . . . . . . . . . . . . . 84
2.9 Robustness checks for overall treatment effects . . . . . . . . . . . . . . . . 87
2.10 Heterogeneity by individual characteristics . . . . . . . . . . . . . . . . . . 91
2.10.1 Policy effect on different education level . . . . . . . . . . . . . . . 92
2.10.2 Policy effect on different gender . . . . . . . . . . . . . . . . . . . . 94
2.11 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102

**3 Absences from school and educational outcomes. Religious observance among ethnic minority students 115**
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116
3.2 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 121
3.2.1 Schooling in England . . . . . . . . . . . . . . . . . . . . . . . . . . 121
3.2.2 Hindu and Muslim festivals . . . . . . . . . . . . . . . . . . . . . . . 122
3.3 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
3.3.1 Data and outcomes . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
3.3.2 Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125
3.3.3 Descriptive statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
3.4 Empirical strategy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132
3.5 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
3.5.1 Religious festivals and absences from school . . . . . . . . . . . . . . 139
3.5.2 Absences and education outcomes . . . . . . . . . . . . . . . . . . . . 139
3.5.3 Heterogeneous effects . . . . . . . . . . . . . . . . . . . . . . . . . . 143
3.6 Robustness checks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 148
3.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 152

References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 157
Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 158

# Chapter 1

# Parental investment and cognitive outcomes in adolescence: evidence from China

**Ziyi Huang***

**Abstract:** *This paper examines how family tutoring and private tutoring affect students' outcomes in China, distinguishing between school test ranks and cognitive test ranks. Using longitudinal data from the China Education Panel Survey (CEPS) and multiple empirical strategies, we find that family tutoring has consistently negative or insignificant effects across outcomes. Private tutoring improves school test rankings, particularly in mathematics, but shows no benefits for cognitive performance. Further analysis indicates that these gains operate through enhanced test-taking ability rather than cognitive skill development: one additional weekly hour of private tutoring raises test-taking ability by 0.005-0.008 standard deviations. The results suggest that tutoring institutions specialize in test-taking ability such as exam-oriented strategies that yield short-term advantages in high-stakes tests but may crowd out development in cognitive ability.*

**Keywords:** Parental inputs, children's educational performance, China

**JEL code:** I20, I21

*Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. Correspondence: zh17565@essex.ac.uk.

## 1.1 Introduction

A large body of research in economics and education has examined how educational investments influence children's academic achievement and skill formation. Classic human capital theories emphasize the role of parental inputs in shaping abilities (Becker, 1962, 1965), and subsequent empirical studies have explored the effects of family involvement and private tutoring on students' academic outcomes (Baker et al., 2001; Buchmann et al., 2010; Byun, 2014; Dang, 2007; Datcher-Loury, 1988; Hill & Stafford, 1974, 1980; Kang & Park, 2021; Loyalka & Zakharov, 2016; Sun et al., 2020; Zhang, 2011, 2013; Zhang & Liu, 2016). Much of the existing literature uses school-based test results as the measure of student performance, or relies on standardised cognitive tests as a proxy. Only a limited number of studies on educational investment explicitly distinguish between these two types of measures and examine them simultaneously (Guo et al., 2020).

The distinction between the two assessments is non-trivial. Standardized cognitive tests are designed to measure fundamental reasoning, problem-solving, and abstract thinking, typically through novel tasks administered in low-stakes settings. These tests assess students' underlying cognitive ability, for which they are neither specifically prepared nor trained. By contrast, school tests are embedded in curriculum-driven and high-stakes institutional contexts (Hanushek & Woessmann, 2008; Heckman & Kautz, 2012), such as the GCSE in England or the Senior High School and College Entrance Examinations in China. These assessments often take students' relative standing within the cohort into account, and progression or selection decisions are not based solely on absolute score thresholds. Consequently, beyond basic subject targets, school tests capture a broader set of competencies which including test-taking strategies, adaptive behaviors, and time management skills. Distinguishing between cognitive ability and test-taking ability is therefore essential for understanding how educational investments translate into academic success. Building on this distinction, this paper asks whether private tutoring and family tutoring affect students' academic performance through their cognitive skills, or whether they primarily cultivate test-taking ability.

This study makes the following contributions. First, much of the existing literature evaluates tutoring using either school-based test results (Buchmann et al., 2010; Dang, 2007; Loyalka & Zakharov, 2016) or standardized cognitive assessments (Baker et al., 2001; Byun, 2014; Kang & Park, 2021), and relatively few studies examine both outcomes simultaneously within the same empirical framework (Guo et al., 2020). By analysing both school examination rankings and cognitive test scores using CEPS data, this paper provides direct evidence on how private tutoring and family tutoring affect these two distinct dimensions of student performance. Second, while previous studies often focus on whether tutoring improves academic performance (Datcher-Loury, 1988; Hill & Stafford, 1974, 1980; Sun et al., 2020; Zhang, 2011, 2013; Zhang & Liu, 2016), they rarely distinguish whether improvements arise from cognitive skill formation or from exam-oriented competencies. This paper addresses this limitation by constructing a proxy measure of test-taking ability, obtained by regressing standardized school test ranks on standardized cognitive test ranks and interpreting the residual component as the test-taking ability. Third, by doing the analysis within China's exam-oriented education system, we provide new evidence on the mechanisms through which educational investments operate. This offers broader insights into how test-based education systems shape learning outcomes.

Following Del Bono et al. (2016), Fiorini and Keane (2014), and Todd and Wolpin (2003, 2007), we employ multiple empirical strategies—including contemporaneous, cumulative, two-way fixed-effect, value-added, and cumulative value-added models—to estimate the effects of tutoring on both school test and cognitive test outcomes. We then treat test-taking ability as a potential channel, allowing us to examine whether tutoring improves school performance through cognitive ability formation, test-taking ability, or both.

Our main findings are as follows. Family tutoring has insignificant or negative effects on students' performance in mathematics, Chinese exams, and cognitive tests across specifications. Private tutoring, by contrast, improves test rankings in mathematics primarily by enhancing test-taking ability. However, it reduces performance on cognitive tests in most models and shows no significant effect in the two-way fixed effects specification. While private tutoring is measured in hours per week and family tutoring captures the frequency of parental involvement—meaning that the estimated coefficients are not directly comparable in magnitude—their signs consistently indicate the direction of the relationships between different inputs and students' outcomes. These results suggest that family tutoring provides limited effective academic support, while private tutoring institutions specialize in developing non-cognitive strategies rather than improving cognitive ability. Furthermore, the emphasis on test-taking ability may even crowd out cognitive development, given the distinct logic underlying the two types of assessments. Thus, although private tutoring can raise test performance in the short run, it may have adverse implications for long-run cognitive ability.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and key variables. Section 4 outlines the empirical strategy. Section 5 presents the results and mechanism analysis. Section 6 provides robustness checks, and Section 7 concludes with a discussion of the findings.

## 1.2 Literature Review

A substantial body of research has examined how family and private tutoring shape student outcomes. A key issue in this literature is how student performance is measured: some studies evaluate school-based test performance, while others focus on standardized cognitive assessments designed to capture underlying abilities. This distinction is important because school tests reflect not only cognitive ability but also exam-oriented competencies, whereas standardized cognitive ability assessments are intended to measure core reasoning and problem-solving skills. Building on this framework, we review international and Chinese evidence before outlining this paper's contribution.

### 1.2.1 International Evidence

International studies suggest that private tutoring is more strongly associated with school tests than with cognitive test outcomes. In Russia, having private tutoring improves Unified State Exam results for high-achieving students (Loyalka & Zakharov, 2016), and in Vietnam, more household tutoring expenditures can raise students' lower-secondary school grades (Dang, 2007). In the United States, SAT preparation courses are positively correlated with scores, although access is strongly stratified by socioeconomic status (Buchmann et al., 2010). By contrast, cross-national evidence from TIMSS finds no significant impact of private tutoring on mathematics cognitive assessments (Baker et al., 2001), and studies from Korea show tutoring only having modest and inconsistent effects on mathematics cognitive scores (Byun, 2014; Kang & Park, 2021). Taken together, these findings indicate that private tutoring primarily improves school test performance, with different benefits across student groups.

Family inputs, in contrast, play a central role in shaping cognitive outcomes. Evidence from the United States shows that maternal time has a causal positive effect on children's cognitive test performance (Carneiro & Rodrigues, 2009; Villena-Roldan & Ríos-Aguilar, 2012). Longitudinal evidence from the UK Millennium Cohort Study also shows persistent positive effects of early maternal investments on cognitive outcomes (Del Bono et al., 2016). Australian panel data indicates that educational activities with parents are the most productive inputs for cognitive skill formation (Fiorini & Keane, 2014), and estimates from the production function show that much of the gap in cognitive outcomes between groups comes from differences in family inputs. (Todd & Wolpin, 2007). Evidence from Denmark further shows that parental developmental can improves school test outcomes, particularly for low-SES children (Thomsen, 2015). Overall, the international evidence highlights a consistent result: private tutoring is more likely to improve performance in school tests, whereas family investments are more likely to benefit cognitive skill development in early childhood.

### 1.2.2 Evidence from China

Chinese studies are broadly consistent with the international literature. Private tutoring can improve school test performance, but the effects vary across groups. Using data from the National College Entrance Examination, Zhang (2011, 2013) find that private tutoring helps low-achieving urban students but harms rural students. Zhang and Liu (2016) show that small-group tutoring raises scores, while large-class tutoring does not. Evidence from CEPS indicates that private tutoring has no overall effect on midterm scores, although it improves English performance for rural students (Sun et al., 2020).

When outcomes are measured using cognitive tests, however, the benefits of private tutoring are much weaker. Using PISA 2015, Liao and Huang (2018) find no significant link between science tutoring and scientific literacy, with small gains only among high-SES students. CEPS-based studies also show that private tutoring produces limited improvements in subject-specific school tests and no significant effects on broader cognitive skills (Guo et al., 2020). By contrast, family inputs—such as parental knowledge and interactive activities—strongly predict gains in children's cognitive and socio-emotional development (Zhong et al., 2020). Consistent with international evidence, Chinese studies confirm that private tutoring is more closely associated with school test performance, whereas family inputs play a greater role in cognitive development.

### 1.2.3 Distinguishing School Test Scores from Cognitive Test Scores

Besides literature mentioned above, some literature formally models the relationship between school tests and cognitive test scores. Hansen et al. (2004) treat both outcomes as functions of a common latent ability and use structural methods to separate the schooling component from the ability component, showing that achievement tests capture both education-driven gains and underlying cognition.

Subsequent work shows that school tests also load on non-cognitive and situational factors. Borghans et al. (2016) find that personal characteristic, especially conscientiousness, explain substantial variation in grades beyond IQ. Gneezy et al. (2019) show that incentives can raise achievement test scores, particularly in low-stakes settings, and Reyes (2023) decompose performance into ability and "cognitive endurance," and cognitive endurance can predict long-run outcomes.

Taken together, this literature indicates that school test results reflect a composite of cognitive skill, effort, and behavioural adaptation under test conditions. Methodologically, these findings motivate approaches that partial out cognitive ability—typically by regressing test scores on cognitive test scores—to isolate a residual interpreted as test-taking ability. This approach underpins our empirical strategy, which uses CEPS data to measure test-taking ability and to examine whether private tutoring enhances cognitive skill, test-taking ability, or both.

### 1.2.4 Contribution

This paper contributes to the literature in following respects. First, motivated by evidence that school test scores capture both cognitive skills and non-cognitive abilities, we construct a proxy for students' test-taking ability by regressing school test performance on cognitive test scores and using the residual as a measure of test-taking ability. Second, we estimate the effects of private tutoring and family tutoring on cognitive skill and test-taking ability, thereby identifying the channels through which different forms of educational investment operate. Third, we apply this approach in China's highly exam-oriented education system. This allows us to provide new evidence on how tutoring shapes student achievement, with implications for other systems where high-stakes tests play an important role.

## 1.3 Data

### 1.3.1 China Education Panel Survey

The China Education Panel Survey (CEPS) is a large-scale, nationally representative longitudinal dataset. The survey employed a sampling strategy based on the average education level of the population and the proportion of the migrated population. Twenty-eight nationwide county-level region were randomly selected. Within these region, 112 schools and 438 classes were randomly chosen, and all students in the selected classes were surveyed. CEPS collects information through five sets of questionnaires covering students, parents, homeroom teachers, subject teachers, and school administrators.

The baseline survey was conducted in the 2013–14 academic year, followed by the second wave in 2014–15, which tracked grade 8 students who had been in grade 7 during the baseline. The first wave surveyed 10,279 students, while the second wave included 10,750 students. From these two waves, we construct a balanced panel sample according to the following steps:

(i) We exclude students who enrolled in schools through non-random channels—such as purchasing school-district housing, relying on personal connections, or providing gifts—in order to mitigate selection bias.[2] A total of 7,277 students remain.

(ii) We exclude schools that did not randomly assign students to classes, thereby avoiding selection bias at the class level. A total of 6,004 students remain.

(iii) Finally, we keep students with complete information on all variables used in the regressions and who responded in both survey waves. This results in 6,838 observations corresponding to 3,419 students.

### 1.3.2 Outcome Variables in CEPS

The CEPS dataset provides two primary measures of educational outcomes. The first is a standardized cognitive test administered through the CEPS questionnaire, and the second is students' school-reported midterm test scores in Mathematics, Chinese, and English.

The CEPS cognitive test consists of four parts, including language, image recognition, calculation, and logic. In wave 1, students answered 20 questions within 15 minutes, whereas in wave 2 they completed 35 questions within 30 minutes. To account for differences in students' prior performance, CEPS employed an ability-adjusted design in wave 2, stratifying students into three groups and administering papers of varying difficulty. To ensure cross-group comparability, CEPS applies a three-parameter logistic Item Response Theory (IRT) model that adjusts for item difficulty, discrimination, and guessing[3].

The second outcome measure is school-based midterm test scores. These tests vary across schools and waves. Most scores range from 0 to 150, and some are scaled from 0 to 100 or 0 to 120. The level of difficulty may also differ across schools, since students' average academic abilities are different in different schools. Therefore, the raw scores from these tests are not directly comparable across schools or waves.

[2]This sample selection is based on the question: "Which of the following did your family do in order to enrol this child in this school?" Answers include "Asking friends for help"; "Giving presents to the related government/school leaders"; "Paying extra fees"; "Buying a house/apartment in the 'education district'"; "Changing the location of the Hukou of your family"; "Transferring the Hukou of your family or this child under other relatives or friends"; "Letting this child take all kinds of achievement tests"; "Other"; "We did none of the above." Students whose families selected any of the first eight options are treated as having enrolled through non-random channels and are excluded from the sample. Students whose families reported "We did none of the above" are treated as having been enrolled through regular admission procedures.

[3]Design details available at: http://ceps.ruc.edu.cn/xmwd/jsbg.htm

To ensure comparability, we convert test scores into school-by-wave ranks. Because our sample is restricted to schools that randomly assign students to classes, we assume that class-level differences within schools are random and therefore do not need additional within-class standardization. We use ranks rather than raw scores because, in the Chinese context, middle school students compete for high school admission at the city level, where relative rather than absolute performance determines placement. e.g., A student is considered to have good performance if she ranks first in her school, regardless of whether her score is 60 or 90.

However, directly using school-by-wave ranks introduces another concern. Ranks are uniformly distributed, whereas students' underlying ability and the effort required to improve performance are not uniformly distributed. As shown in Figure 1.1, both school test scores and cognitive test scores approximately follow a normal distribution. Then, the effort needed for a student in second place to move to first is substantially greater than that required for a student ranked eleventh to move to tenth in a class of twenty. In this case, we apply the rank-based inverse normal transformation[4] (Soloman & Sawilowsky, 2009) to both school test scores and cognitive test scores. After transformation, as shown in Figure 1.1, outcomes are closer to the normal distribution, $N(0,1)$.

### 1.3.3 Test-taking ability

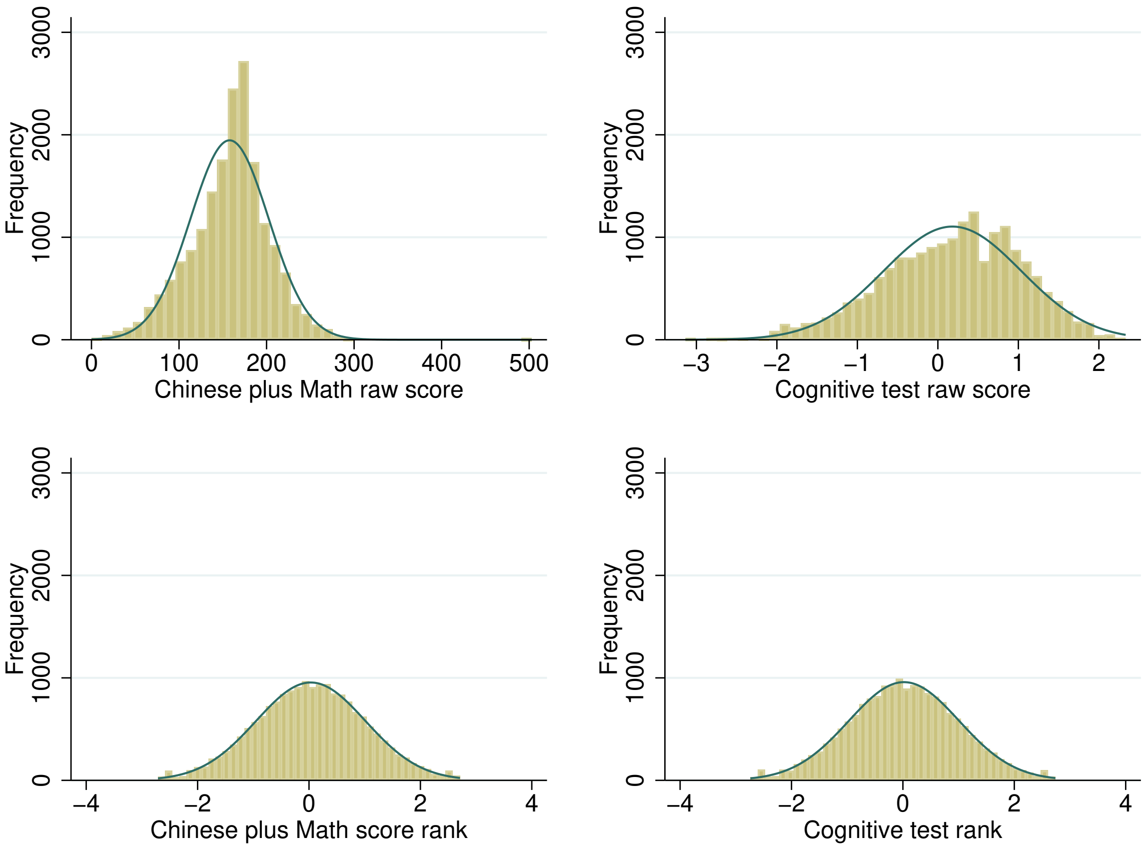
Literature shows school test outcomes can reflect a composite of cognitive skills, effort, and behavioral adaptation under test conditions. In this case, if tutoring is targeting increasing students' school test results, then they are very likely to provide some training on test-taking strategies, adaptive behaviors, and time management. In order to study whether tutoring affects students' school test outcomes through cognitive ability or test-taking ability, we construct a proxy for test-taking ability by isolating the cognitive ability from school test performance.

Operationally, we use the standardized outcome mentioned above and regress school test ranks on cognitive test ranks, controlling for school-by-wave fixed effect[5]. The school-by-wave

[4] The rank-based inverse normal transformation maps the empirical rank $r_i$ of observation $i$ among $n$ observations to the corresponding quantile of the standard normal distribution: $z_i = \Phi^{-1}\left(\frac{r_i - 0.5}{n}\right)$, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard normal distribution. This procedure yields variables approximately distributed as $N(0,1)$ and mitigates the influence of outliers.

[5] We can't add student fixed effect here, since student FE will make the mean of residuals equal to zero within student.

**Figure 1.1:** Distribution of outcome variables




**Note:**Data from CEPS 2013–2015. Outcome variables are shown prior to sample selection and have been transformed using the rank-based inverse normal transformation.

fixed effects absorb variation arising from differences in school quality, grading standards, and exam difficulty over time.[6] The remaining variation therefore captures all factors unrelated to school, year and cognitive ability, including students' personalities, exam effort, time management, short-term motivation, adaptive responses to testing conditions, and random factors. We interpret this residual component as a proxy for students' test-taking ability.

Since the CEPS cognitive assessment consists of four domains: language, image recognition, calculation, and logic. To ensure comparability, we use school test outcomes in Chinese and mathematics as the corresponding measure of school test performance. The estimating equation is:

$$Chinese\,plus\,math\,rank_{its} = \gamma Cognitive\,rank_{it} + school\,wave_{ts} + resid_{its} \tag{1.1}$$

where $Chinese\,plus\,math\,rank_{its}$ is student $i$'s rank of Chinese plus Math in wave $t$ school $s$. $Cognitive\,rank_{it}$ is students' rank of cognitive test in wave $t$ school $s$. $school\,wave_{ts}$ absorbs school-by-wave common shocks. The residual component $resid_{its}$ is interpreted as proxy for test-taking ability. Figure 1.2 presents the scatter plots for this equation.

In Figure 1.2, cognitive rank is positively associated with students' school test outcomes, with an R-squared of 0.252. The residuals, which we interpret as a proxy for test-taking ability, are positive when they lie above the line of best fit and negative when they lie below it. Figure 1.3 then presents the distribution of test-taking ability and its corresponding rank, both of which follow an approximately normal distribution.

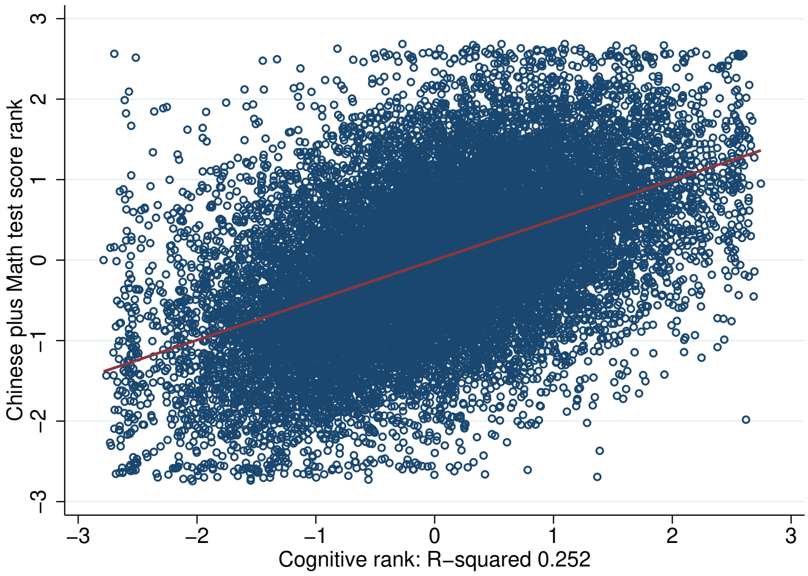
### 1.3.4 Private Tutoring (PT)

The CEPS dataset provides detailed information on private tutoring participation. In wave 1, students directly reported hours and minutes of participation. In wave 2, participation was reported separately for weekdays and weekends using categorical scales. Weekday responses ranged from "None" to "More than four hours," while weekend responses ranged from "None" to "More than eight hours."[7]

[6] The school-by-wave fixed effects are expected to have only a minor impact here, as the outcomes are already standardized within each school and wave.

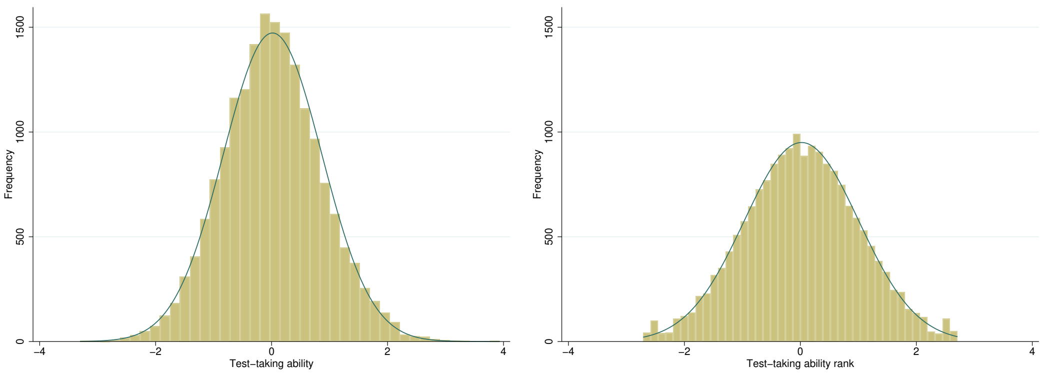
[7] Although the scale remains per day, weekend bands are wider to reflect longer available time.

**Figure 1.2:** Scatter of cognitive test rank and school test rank




**Note:** Data from CEPS 2013-2015.Students with non-missing outcome variables are included. The school-wave fixed effect has been subtracted in the plot. This is the scatter point graph for equation(1).

**Figure 1.3:** Distribution of test-taking ability




**Note:** Data from CEPS 2013-2015.Outcome variables are shown prior to sample selection and have been transformed using the rank-based inverse normal transformation.

Private tutoring is further classified into academic PT (school-subject related) and hobby PT (non-school subjects). Our analysis focuses on academic PT. Categorical responses are converted into continuous measures by taking midpoints of each interval. For weekdays, the coding is: None = 0, Less than 1 hour = 0.5, One to two hours = 1.5, Two to three hours = 2.5, Three to four hours = 3.5, More than four hours = 4.5. For weekends: None = 0, Less than 2 hours = 1, About 2–4 hours = 3, About 4–6 hours = 5, About 6–8 hours = 6, More than 8 hours = 9.

We sum weekday and weekend tutoring hours to obtain total weekly tutoring time, and treat all private tutoring hours as continuous variables. Therefore, the estimated coefficients on private tutoring can be interpreted as the change in the outcome variable associated with an additional hour of tutoring per week. Since the dependent variables are standardized using a rank-based inverse normal transformation, the coefficients represent changes in the outcomes measured in standard deviation units.

### 1.3.5 Family Tutoring (FT)

Unlike asking range of hours in private tutoring, the CEPS questionnaire measures family tutoring through two questions on how often parents (a) check students' homework and (b) provide instructions. Responses are reported on a four-point scale ranging from "Never" to "Almost every day" (None = 1; One or two days per week = 2; Three or four days per week = 3; Almost every day = 4). We transform these categorical responses into continuous days-per-week measures based on the midpoint of each interval (None = 0; One or two = 1.5; Three or four = 3.5; Almost every day = 6). Our FT frequency measure is defined as the average of the two items.

It is important to note that the measurement of family tutoring differs from that of private tutoring. While private tutoring is measured in hours per week, family tutoring captures the frequency of parental involvement in checking homework and providing instruction, where one unit represents doing so on an additional day per week. As a result, the coefficients on FT and PT should not be interpreted as directly comparable in magnitude. Instead, they should be interpreted as indicating the direction and statistical significance of the relationships between each type of parental input and students' outcomes.

### 1.3.6 Control Variables

We include a set of household-level and school-level control variables in the regressions. At the household level, controls include students' *hukou* type, health status, co-residence with father and mother, boarding status, number of siblings, and parents' education and occupation, as well as family economic status. *Hukou* type is categorized as either rural or urban. Health status is a binary indicator of whether the student reports poor health. Parental education is originally reported on a nine-point scale in CEPS, which we recode into a five-point scale for analytical clarity. Specifically, "Technical secondary school or technical school degree," "Vocational high school degree," and "Senior high school degree" are grouped into "High school," while "Junior college degree," "Bachelor degree," and "Master degree or higher" are grouped into "College or above." Parental occupation is recorded in eight categories: unemployed, manager, professional worker, high-skill worker, ordinary worker, self-employed, and other. Although CEPS does not collect parents' incomes directly, the parent questionnaire asks respondents to assess their household's relative economic status on a five-point scale ranging from 1 ("Very low") to 5 ("Very high").

At the school level, we control for class size, teachers' education, and teaching experience. School inputs are widely recognized as critical determinants of student performance, both directly through instructional quality and indirectly through parental educational investments. Teachers' education is coded on a dummy about whether have regular Bachelor or higher degree. Teachers' experience are counting in year, and class size is the number of students in the class.

### 1.3.7 Descriptive statistics

Table 1.1 presents summary statistics for the main variables, disaggregated by gender and by hukou type. The differences before and after sample selection involve school test ranks, cognitive ranks, and the proportion of students with urban *hukou*. After excluding students who were admitted to schools through non-random channels, the mean values of both school test ranks and cognitive ranks increased. This pattern arises because parents who purchase school-district housing typically target higher-quality schools. As a result, their children tend to enrol in schools with higher average performance levels, leading to relatively lower own within-school ranks. Consequently, removing these students raises the average rank in the remaining sample.

Another key difference is that a considerable number of students with rural *hukou* are excluded, primarily because the analysis is restricted to students with complete information across all variables. Rural students are less likely to provide full information, and the small shifts observed in family economic status, parental education, and occupation largely driven by this change in the composition of rural students.

For outcomes, girls rank significantly higher than boys in Chinese and English, by 0.62 and 0.57 standard deviations, respectively, whereas boys perform slightly better in mathematics (0.14 standard deviations). Cognitive test rankings do not differ significantly between boys and girls, but girls have higher rank in test-taking ability rank. Comparing *hukou* groups, urban students rank higher than rural students in English (0.06 standard deviations), indicating that urban–rural disparities persist, although the magnitude of the gap is modest.

Tutoring inputs also vary substantially across groups. Urban students receive on average 5.3 hours of private tutoring (PT) per week, compared to only 2.3 hours for rural students. The urban–rural gap of approximately 3 hours per week is both large and statistically significant. Family tutoring (FT), measured as the frequency of parental help with homework, also shows a rural disadvantage: urban students receive on average 2.5 times of FT per week, whereas rural students receive 2.0 times. Gender differences are less pronounced: boys receive 0.18 more sessions of family tutoring per week than girls.

Family background characteristics reinforce these disparities. Urban students are more likely to have parents with higher education: 34 percent of urban fathers and 28 percent of urban mothers hold a college degree or above, compared with 9 percent and 6 percent among rural parents. Parental occupations also differ markedly, with rural parents concentrated in ordinary labor and self-employment, while urban parents are more frequently employed in professional or managerial positions. Reported family economic status is significantly higher in urban households, with nearly half reporting above-average status, compared to less than one-fifth among rural households. In contrast, gender differences in parental education, occupation, and family economic status are generally small and statistically insignificant.

Overall, the descriptive evidence highlights clear gender differences in subject-specific rankings and substantial urban–rural disparities in both tutoring investments and household background conditions.

**Table 1.1:** Descriptive statistics

| Variable | (1) All samples before selection | (2) All samples after selection | (3) Female | (4) Male | (5) Difference between female and male | (6) Rural hukou | (7) Urban hukou | (8) Difference between rural and urban hukou |
|---|---|---|---|---|---|---|---|---|
| **Outcomes:** | | | | | | | | |
| Chinese score rank | -0.001 | 0.107 | 0.402 | -0.216 | 0.618*** | 0.089 | 0.123 | -0.034 |
| | (0.990) | (0.964) | (0.895) | (0.933) | | (0.948) | (0.979) | |
| Math score rank | -0.001 | 0.091 | 0.155 | 0.020 | 0.135*** | 0.089 | 0.092 | -0.003 |
| | (0.987) | (0.978) | (0.947) | (1.005) | | (0.968) | (0.987) | |
| English score rank | -0.001 | 0.098 | 0.372 | -0.201 | 0.573*** | 0.067 | 0.128 | -0.061*** |
| | (0.989) | (0.971) | (0.898) | (0.960) | | (0.958) | (0.983) | |
| CEPS cognitive test score rank | 0.000 | 0.077 | 0.087 | 0.067 | 0.020 | 0.065 | 0.089 | -0.023 |
| | (0.992) | (0.989) | (0.964) | (1.016) | | (0.972) | (1.006) | |
| Test-taking ability rank | 0.000 | 0.076 | 0.267 | -0.132 | 0.399*** | 0.069 | 0.083 | -0.014 |
| | (0.993) | (0.976) | (0.933) | (0.980) | | (0.959) | (0.993) | |
| **Explanatory variables:** | | | | | | | | |
| Private tutoring: hours/week | 3.864 | 3.815 | 3.782 | 3.851 | -0.069 | 2.268 | 5.306 | -3.039*** |
| | (7.652) | (7.190) | (7.039) | (7.352) | | (5.871) | (7.986) | |
| Family tutoring: times/week | 2.201 | 2.249 | 2.162 | 2.345 | -0.183*** | 2.000 | 2.490 | -0.490*** |
| | (1.622) | (1.648) | (1.641) | (1.651) | | (1.571) | (1.685) | |
| **Time-invariant controls:** | | | | | | | | |
| Gender (male = 1) | 0.527 | 0.479 | 0 | 1 | NA | 0.483 | 0.474 | 0.008 |
| | (0.499) | (0.500) | | | | (0.500) | (0.499) | |
| Age (in month) | 165.261 | 164.287 | 163.785 | 164.834 | -1.050*** | 165.519 | 163.100 | 2.419*** |
| | (10.179) | (9.431) | (9.321) | (9.519) | | (10.236) | (8.415) | |
| Father education level: | | | | | | | | |
| no schooling | 0.008 | 0.006 | 0.004 | 0.007 | -0.002 | 0.009 | 0.003 | 0.006*** |
| | (0.088) | (0.074) | (0.067) | (0.082) | | (0.093) | (0.051) | |
| primary school | 0.138 | 0.127 | 0.126 | 0.128 | -0.001 | 0.201 | 0.056 | 0.145*** |
| | (0.345) | (0.333) | (0.332) | (0.334) | | (0.401) | (0.230) | |
| middle school | 0.502 | 0.491 | 0.500 | 0.481 | 0.019 | 0.605 | 0.380 | 0.224*** |
| | (0.500) | (0.500) | (0.500) | (0.500) | | (0.489) | (0.486) | |
| high school | 0.188 | 0.181 | 0.174 | 0.188 | -0.014 | 0.149 | 0.211 | -0.063*** |
| | (0.391) | (0.385) | (0.379) | (0.391) | | (0.356) | (0.408) | |
| college or higher | 0.165 | 0.196 | 0.195 | 0.197 | -0.002 | 0.037 | 0.349 | -0.312*** |
| | (0.371) | (0.397) | (0.396) | (0.398) | | (0.189) | (0.477) | |
| Mother education level: | | | | | | | | |
| no schooling | 0.036 | 0.029 | 0.026 | 0.032 | -0.007* | 0.044 | 0.014 | 0.031*** |
| | (0.185) | (0.167) | (0.158) | (0.177) | | (0.206) | (0.117) | |
| primary school | 0.191 | 0.176 | 0.172 | 0.181 | -0.009 | 0.276 | 0.080 | 0.196*** |
| | (0.393) | (0.381) | (0.377) | (0.385) | | (0.447) | (0.271) | |
| middle school | 0.477 | 0.473 | 0.483 | 0.463 | 0.020* | 0.543 | 0.406 | 0.138*** |
| | (0.499) | (0.499) | (0.500) | (0.499) | | (0.498) | (0.491) | |
| high school | 0.154 | 0.147 | 0.142 | 0.152 | -0.011 | 0.104 | 0.188 | -0.084*** |
| | (0.361) | (0.354) | (0.349) | (0.359) | | (0.305) | (0.391) | |
| college or higher | 0.143 | 0.175 | 0.178 | 0.172 | 0.006 | 0.033 | 0.313 | -0.280*** |
| | (0.350) | (0.380) | (0.383) | (0.377) | | (0.177) | (0.464) | |
| N | 21,029 | 6,838 | 3,566 | 3,272 | | 3,356 | 3,482 | |

**Note:** Data from CEPS 2013-2015. All observations prior to sample selection, including all raw data, constitute the raw sample. The total number of observations therefore reflects the size of the raw dataset, although the extent of missing values varies across variables. All samples after selection are observations in the main estimations. Test scores ranks and cognitive outcomes rank is calculated within wave and school. The rank-based inverse normal transformation makes outcomes close to normal distribution but not strictly follow N(0,1), and it will follow N(0,1) when sample size close to infinity. All outcomes are calculated after sample selection.

Table 1.1b: Descriptive statistics: continued

| Variable | (1) All samples before selection | (2) All samples after selection | (3) Female | (4) Male | (5) Difference between female and male | (6) Rural hukou | (7) Urban hukou | (8) Difference between rural and urban hukou |
|---|---|---|---|---|---|---|---|---|
| **Time-variant controls:** | | | | | | | | |
| Urban hukou | 0.454 | 0.509 | 0.513 | 0.505 | 0.008 | 0 | 1 | NA |
| | (0.498) | (0.500) | (0.500) | (0.500) | | | | |
| Healthy (good = 1) | 0.969 | 0.976 | 0.975 | 0.976 | -0.001 | 0.968 | 0.983 | -0.015*** |
| | (0.174) | (0.154) | (0.156) | (0.153) | | (0.176) | (0.130) | |
| Not live with mother | 0.131 | 0.116 | 0.110 | 0.123 | -0.014* | 0.145 | 0.088 | 0.057*** |
| | (0.337) | (0.321) | (0.313) | (0.329) | | (0.353) | (0.284) | |
| Not live with father | 0.178 | 0.155 | 0.152 | 0.159 | -0.008 | 0.189 | 0.123 | 0.067*** |
| | (0.383) | (0.362) | (0.359) | (0.366) | | (0.392) | (0.328) | |
| Boarding at school | 0.308 | 0.221 | 0.218 | 0.224 | -0.005 | 0.364 | 0.083 | 0.281*** |
| | (0.462) | (0.415) | (0.413) | (0.417) | | (0.481) | (0.276) | |
| Number of siblings | 0.702 | 0.661 | 0.731 | 0.586 | 0.145*** | 0.922 | 0.410 | 0.512*** |
| | (0.859) | (0.812) | (0.865) | (0.742) | | (0.802) | (0.738) | |
| Economic status: | | | | | | | | |
| very poor | 0.040 | 0.032 | 0.030 | 0.035 | -0.005 | 0.049 | 0.016 | 0.032*** |
| | (0.196) | (0.177) | (0.170) | (0.184) | | (0.216) | (0.127) | |
| somewhat poor | 0.175 | 0.150 | 0.144 | 0.155 | -0.011 | 0.206 | 0.095 | 0.112*** |
| | (0.380) | (0.357) | (0.352) | (0.362) | | (0.405) | (0.293) | |
| moderate | 0.723 | 0.746 | 0.756 | 0.735 | 0.021** | 0.690 | 0.800 | -0.109*** |
| | (0.447) | (0.435) | (0.429) | (0.442) | | (0.463) | (0.400) | |
| somewhat rich | 0.058 | 0.069 | 0.067 | 0.072 | -0.005 | 0.053 | 0.085 | -0.033*** |
| | (0.234) | (0.254) | (0.250) | (0.258) | | (0.224) | (0.279) | |
| very rich | 0.004 | 0.003 | 0.003 | 0.003 | 0.000 | 0.002 | 0.004 | -0.002* |
| | (0.062) | (0.054) | (0.053) | (0.055) | | (0.042) | (0.063) | |
| Father's occupation: | | | | | | | | |
| unemployed | 0.028 | 0.031 | 0.030 | 0.031 | -0.001 | 0.020 | 0.041 | -0.021*** |
| | (0.165) | (0.172) | (0.171) | (0.173) | | (0.139) | (0.199) | |
| manager | 0.128 | 0.137 | 0.128 | 0.147 | -0.019** | 0.044 | 0.227 | -0.182*** |
| | (0.334) | (0.344) | (0.335) | (0.354) | | (0.206) | (0.419) | |
| professional labour | 0.058 | 0.058 | 0.064 | 0.051 | 0.013** | 0.031 | 0.084 | -0.053*** |
| | (0.235) | (0.234) | (0.246) | (0.221) | | (0.173) | (0.278) | |
| high-skill labour | 0.219 | 0.225 | 0.214 | 0.237 | -0.023** | 0.211 | 0.239 | -0.028*** |
| | (0.413) | (0.418) | (0.410) | (0.425) | | (0.408) | (0.427) | |
| ordinary labour | 0.164 | 0.163 | 0.159 | 0.167 | -0.008 | 0.191 | 0.135 | 0.056*** |
| | (0.370) | (0.369) | (0.365) | (0.373) | | (0.393) | (0.342) | |
| self-employed | 0.171 | 0.173 | 0.182 | 0.164 | 0.018* | 0.178 | 0.169 | 0.010 |
| | (0.376) | (0.379) | (0.386) | (0.370) | | (0.383) | (0.374) | |
| farmer | 0.183 | 0.167 | 0.174 | 0.160 | 0.014 | 0.287 | 0.051 | 0.236*** |
| | (0.387) | (0.373) | (0.379) | (0.366) | | (0.452) | (0.220) | |
| other | 0.049 | 0.046 | 0.048 | 0.043 | 0.005 | 0.038 | 0.054 | -0.016*** |
| | (0.217) | (0.209) | (0.214) | (0.203) | | (0.190) | (0.226) | |
| N | 21,029 | 6,838 | 3,566 | 3,272 | | 3,356 | 3,482 | |

**Note:** Data from CEPS 2013-2015. All observations prior to sample selection, including all raw data, constitute the raw sample. The total number of observations therefore reflects the size of the raw dataset, although the extent of missing values varies across variables. All samples after selection are observations in the main estimations.

Table 1.1c: Descriptive statistics: continued

| Variable | (1) All samples before selection | (2) All samples after selection | (3) Female | (4) Male | (5) Difference between female and male | (6) Rural hukou | (7) Urban hukou | (8) Difference between rural and urban hukou |
|---|---|---|---|---|---|---|---|---|
| Mother's occupation: | | | | | | | | |
| unemployed | 0.106 | 0.111 | 0.115 | 0.107 | 0.007 | 0.096 | 0.126 | -0.030*** |
| | (0.307) | (0.314) | (0.319) | (0.309) | | (0.295) | (0.332) | |
| manager | 0.075 | 0.081 | 0.082 | 0.079 | 0.002 | 0.021 | 0.138 | -0.116*** |
| | (0.264) | (0.272) | (0.274) | (0.271) | | (0.145) | (0.345) | |
| professional labour | 0.049 | 0.055 | 0.056 | 0.054 | 0.002 | 0.018 | 0.091 | -0.073*** |
| | (0.216) | (0.229) | (0.231) | (0.227) | | (0.135) | (0.288) | |
| high-skill labour | 0.126 | 0.137 | 0.132 | 0.143 | -0.012 | 0.094 | 0.179 | -0.084*** |
| | (0.332) | (0.344) | (0.338) | (0.351) | | (0.293) | (0.383) | |
| ordinary labour | 0.223 | 0.228 | 0.210 | 0.248 | -0.037*** | 0.250 | 0.207 | 0.043*** |
| | (0.416) | (0.420) | (0.408) | (0.432) | | (0.433) | (0.405) | |
| self-employed | 0.157 | 0.153 | 0.159 | 0.145 | 0.014 | 0.154 | 0.151 | 0.003 |
| | (0.364) | (0.360) | (0.366) | (0.353) | | (0.361) | (0.358) | |
| farmer | 0.211 | 0.185 | 0.192 | 0.176 | 0.016* | 0.320 | 0.055 | 0.265*** |
| | (0.408) | (0.388) | (0.394) | (0.381) | | (0.466) | (0.227) | |
| other | 0.053 | 0.050 | 0.053 | 0.046 | 0.007 | 0.046 | 0.054 | -0.008 |
| | (0.224) | (0.218) | (0.225) | (0.210) | | (0.209) | (0.226) | |
| Teacher education level: | | | | | | | | |
| (whether have regular Bachelor or higher) | | | | | | | | |
| Teacher of Chinese | 0.444 | 0.466 | 0.472 | 0.460 | 0.012 | 0.409 | 0.522 | -0.112*** |
| | (0.497) | (0.499) | (0.499) | (0.498) | | (0.492) | (0.500) | |
| Teacher of Math | 0.485 | 0.486 | 0.491 | 0.481 | 0.009 | 0.365 | 0.603 | -0.238*** |
| | (0.500) | (0.500) | (0.500) | (0.500) | | (0.482) | (0.489) | |
| Teacher of English | 0.445 | 0.442 | 0.455 | 0.428 | 0.026** | 0.416 | 0.467 | -0.050*** |
| | (0.497) | (0.497) | (0.498) | (0.495) | | (0.493) | (0.499) | |
| Teacher teaching experience: years | | | | | | | | |
| Teacher of Chinese | 15.987 | 15.249 | 15.306 | 15.188 | 0.118 | 14.799 | 15.684 | -0.886*** |
| | (9.101) | (8.983) | (8.848) | (9.129) | | (9.219) | (8.729) | |
| Teacher of Math | 16.696 | 17.319 | 17.197 | 17.452 | -0.255 | 17.565 | 17.082 | 0.483** |
| | (8.701) | (8.441) | (8.456) | (8.423) | | (8.309) | (8.561) | |
| Teacher of English | 15.347 | 15.534 | 15.575 | 15.490 | 0.086 | 15.042 | 16.009 | -0.967*** |
| | (9.069) | (9.008) | (9.007) | (9.011) | | (9.433) | (8.553) | |
| Number of students in class | 51.035 | 49.763 | 49.646 | 49.890 | -0.244 | 49.890 | 49.640 | 0.251 |
| | (13.396) | (13.182) | (12.883) | (13.500) | | (13.687) | (12.676) | |
| N | 21,029 | 6,838 | 3,566 | 3,272 | | 3,356 | 3,482 | |

**Note:** Data from CEPS 2013-2015. All observations prior to sample selection, including all raw data, constitute the raw sample. The total number of observations therefore reflects the size of the raw dataset, although the extent of missing values varies across variables. All samples after selection are observations in the main estimations.

## 1.4 Methods

We frame our analysis within the education production function defined by Hanushek (2020), which conceptualizes student achievement as the outcome of inputs from families, peers, and schools.

In estimating this relationship, three potential sources of endogeneity must be addressed. First, omitted variable bias may arise if unobserved factors such as students' innate ability simultaneously affect both educational outcomes and the likelihood of receiving tutoring. Second, reverse causality is possible: while parental inputs may improve children's outcomes, parents may also increase inputs in response to poor performance. Third, self-selection in schools or classrooms creates additional bias. Parents with the ability to choose schools or classrooms for their children are likely to select environments and peers, introducing correlation between unobserved peer quality and parental inputs.

Previous studies have attempted to address these concerns by using peer behaviour as an instrumental variable (IV). For example, Yang and Zhao (2020) exploit parents in other classes within the same school as instruments for parenting styles, and for robustness they additionally use the proportion of other parents' reported parenting styles within the same class. However, this approach may not satisfy the exclusion restriction, as parental behaviour within the same survey wave can directly influence other parents, generating spillovers that affect both the peer and the parent herself. Similarly, studies on private tutoring in China (Yang & Zhao, 2020; Zhang, 2013) have also relied on peer effects as instruments. However, peer tutoring is unlikely to be a valid IV, since peers' tutoring not only influences student's probability of receiving tutoring, but also directly affects the peers' academic outcomes, which in turn can influence the student through other channels. This violates the exclusion restriction and leads to biased IV estimates.

An alternative approach is the cumulative value-added instrumental variables (CVA-IV) model, widely applied in panel data settings (Andrabi et al., 2011; Del Bono et al., 2016). This method uses two-wave lagged outcomes as the instrument for lagged outcomes. However, we can only use CVA model in the analysis since CEPS contains only two waves.

The remainder of this section introduces the econometric models employed, each relying on different identifying assumptions. Following Del Boca et al. (2012), Fiorini and Keane (2014), and Todd and Wolpin (2003, 2007), we implement a series of specifications designed to address particular biases.

Formally, we write children's education production function as:

$$Y_{it} = FT'_{i\{1\times t\}}\alpha_{\{1\times t\}} + PT'_{i\{1\times t\}}\beta_{\{1\times t\}} + X'_{i\{G\times t\}}\delta_{\{G\times t\}} + \eta_i + e_{it} \tag{1.2}$$

where $Y_{it}$ denotes the outcome of student $i$ at time $t$, $FT_{it}$ and $PT_{it}$ represent family tutoring and private tutoring inputs, $X_{it}$ is a vector of background characteristics, $\eta_i$ is an unobserved time-invariant individual effect, and $e_{it}$ is a time-varying error term.

Our analysis relies on the key assumption that family tutoring and private tutoring enter the production function additively and are separable.

### 1.4.1 Contemporaneous Model

We begin by estimating a contemporaneous specification using ordinary least squares (OLS). This model assumes that students' outcomes in a given wave respond to tutoring received during the same wave. It relies exclusively on current-wave variation in tutoring and is formally specified as:

$$Y_{it} = FT'_{it}\alpha_t + PT'_{it}\beta_t + X'_{it\{G\}}\delta_{t\{G\}} + e_{it} \tag{1.3}$$

where $Y_{it}$ denotes student $i$'s performance in wave $t$, $FT_{it}$ and $PT_{it}$ represent family and private tutoring inputs, and $X_{it}$ is a vector of $G$ control variables. The terms $e_{it}$ capture unobserved heterogeneity and idiosyncratic shocks.

The contemporaneous model rests on two key assumptions:

- Student performance depends only on current inputs.
- The included control variables sufficiently proxy for omitted variables so that inputs are uncorrelated with the error term.

OLS estimation is the most widely applied method in empirical economics. It is intuitive, easy to implement, and provides a useful benchmark for preliminary analysis. However, the assumptions underlying the contemporaneous model are restrictive. If reverse causality exists, such that parents allocate more inputs to children with lower performance, the estimates are biased downward. If higher-performing students attract greater parental inputs, the estimates are biased upward.

### 1.4.2 Two-Way Fixed Effects Model (TWFE)

Two-way fixed effects (TWFE) model is widely used in longitudinal analyses with repeated observations on the same individuals, as it allows us to control for both time-invariant individual heterogeneity and common shocks at the wave level. Intuitively, the TWFE model relates within-individual deviations from their own mean outcomes to within-individual deviations from their own mean inputs, netting out wave-specific averages. The estimating equation is:

$$Y_{it} = FT'_{it}\alpha_t + PT'_{it}\beta_t + X'_{it\{G\}}\delta_{t\{G\}} + \tau_t + \omega_i + e_{it} \tag{1.4}$$

Where $\tau_t$ is the wave fixed effect and $\omega_i$ is the individual fixed effect. The TWFE model relies on the following assumptions:

- Changes in student performance are explained by changes in inputs after controlling for both time-invariant individual heterogeneity and common shocks at the wave level.
- The effects of both observed and unobserved variables that are absorbed by fixed effects are constant across individuals and over time.
- Inputs are strictly exogenous with respect to the idiosyncratic error term, conditional on both individual and time fixed effects.

The TWFE model identifies how within-individual changes in inputs, net of time-specific shocks, translate into changes in outcomes. Compared with specifications without time fixed effects, it has the advantage of eliminating bias from both time-invariant unobserved heterogeneity and period-specific common shocks.

Nevertheless, identification relies only on individuals who vary their inputs across waves, and the model assumes that the parameters of the education production function remain constant across periods. If parameters differ across waves, the resulting bias may be either upward or downward depending on the direction of the slope change. Additional bias may also arise if time-varying unobserved factors at the individual level are correlated with inputs, which are not fully captured by the fixed effects.

### 1.4.3 Value-Added Model (VA)

The value-added (VA) model extends the contemporaneous specification by including the lagged outcome as a regressor. This approach assumes that current-wave inputs affect changes in students' outcomes relative to their prior level, thereby accounting for baseline differences in achievement. The estimating equation is:

$$Y_{it} = FT'_{it}\alpha_t + PT'_{it}\beta_t + X'_{it\{G\}}\delta_{t\{G\}} + \lambda_t Y_{i,t-1} + e_{it} \tag{1.5}$$

The key assumptions underlying this model are:

- The lagged outcome serves as an effective control for past inputs and prior information.
- The effect of inputs declines with age at a rate captured by $\lambda_t$.

The VA model links current performance with prior performance. The inclusion of the lagged outcome captures unobserved abilities and proxies for omitted inputs that influence outcomes before the lagged test. Relative to the contemporaneous model, this specification helps to mitigate bias from reverse causality, since current inputs are often driven by lagged outcomes, and it better controls for unobserved heterogeneity.

However, the VA model also has limitations. It does not fully address simultaneity: if parental investments respond to new information about ability or performance that emerges between waves, reverse causality may still arise. For example, changes in learning materials across terms may generate differences in ability that are not captured by the lagged measure. Moreover, lagged outcomes are subject to measurement error,[8] and ability may not operate solely through a one-time endowment.[9] As a result, the direction of bias is difficult to predict, since measurement error in the lagged outcome can contaminate the entire vector of estimated coefficients.

### 1.4.4 Cumulative model (CU)

The cumulative (CU) specification extends the OLS model by incorporating lagged inputs as additional regressors. This approach assumes that students' outcomes are jointly determined by contemporaneous inputs and the cumulative effects of past inputs. The estimating equation is:

$$Y_{it} = FT'_{i\{1\times t\}}\alpha_{\{1\times t\}} + PT'_{i\{1\times t\}}\beta_{\{1\times t\}} + X'_{i\{G\times t\}}\delta_{\{G\times t\}} + e_{it}{}^{10} \tag{1.6}$$

The main assumptions behind this model are:

- Performance relate to current and previous inputs.
- The control variables proxy for all omitted variables.

Unlike the value-added model, which includes the lagged outcome to absorb unobserved persistent factors, the CU model uses lagged inputs directly to capture the cumulative effects of past investments on current outcomes.

### 1.4.5 Cumulative Value-Added Model (CVA)

The cumulative value-added (CVA) model combines elements of the cumulative and value-added specifications. It assumes that changes in students' outcomes are jointly determined by contemporaneous inputs and the cumulative effects of past inputs, while controlling for prior achievement through the lagged outcome. The estimating equation is:

[8] Test scores are inherently noisy measures of latent achievement.

[9] This implies that the lagged outcome cannot fully capture abilities that affect performance between two test waves (Andrabi et al., 2011).

[10] In this chapter, CU model only contains two periods.e.g. $t = 2$

$$Y_{it} = FT'_{i\{1\times t\}}\alpha_{\{1\times t\}} + PT'_{i\{1\times t\}}\beta_{\{1\times t\}} + X'_{i\{G\times t\}}\delta_{\{G\times t\}} + \lambda_t Y_{i,t-1} + e_{it} \tag{1.7}$$

This model is based on the following assumptions:

- Student performance is determined by both current and lagged inputs.
- The effect of inputs declines with age at a rate captured by $\lambda_t$.
- Lagged outcomes and lagged inputs jointly provide effective controls for prior information and unobserved heterogeneity.

In the CVA framework, the lagged outcome captures unobserved abilities and omitted inputs, while the inclusion of lagged inputs helps mitigate reverse causality. Compared with the contemporaneous, cumulative, and value-added specifications, the CVA model addresses a broader set of endogeneity concerns and is generally less biased.

Nonetheless, the model has important limitations. It does not fully eliminate reverse causality within the current wave, since parents may still adjust inputs in response to newly observed information between tests. Moreover, the lagged outcome regressor may be contaminated by measurement error,[11] and ability may influence performance through channels other than an initial one-time endowment, leading to residual bias.

## 1.5 Estimation Results

In this section, we present estimation results based on the models introduced above. We begin by examining the effect of private tutoring and family tutoring on students' school test ranks. Then we describe how tutoring affects students' cognitive ability and test-taking ability differently.

Table 1.2 reports the estimated effects of tutoring on students' school test performance, and Table 1.3 shows the results of tutoring on students' cognitive ability and test-taking ability. In both Table 1.2 and Table 1.3, columns (1) present contemporaneous effects. Column (2) introduces two-way fixed effects (TWFE), capturing how within-student and within-time changes

[11] Test scores are noisy proxies for latent achievement.

in tutoring are associated with changes in performance. Column (3) reports the value-added specification, which includes both contemporaneous tutoring inputs and the lagged dependent variable. Column (4) extends the analysis using the cumulative model, which augments the contemporaneous specification with observable lagged inputs. Finally, column (5) combines both approaches in a cumulative value-added specification, incorporating lagged inputs alongside the lagged outcome.

### 1.5.1 School Test Rank

In this part, we show effect of private tutoring and family tutoring on students' Chinese, Math, and English test score rank separately in different panels. We also present effect of tutoring on students rank after adding Chinese and math scores, since this rank is used in estimating test-taking ability.

For all subjects' midterm ranks and Chinese plus math rank, family tutoring (FT) consistently shows negative contemporaneous associations, with coefficients around $-0.02$ to $-0.06$ standard deviations. However, these effects largely vanish once student fixed effects are introduced, suggesting that part of the negative association reflects time-invariant unobserved heterogeneity. In other words, family tutoring is more likely to be undertaken by students with persistently weaker academic performance. For example, in Panel A column (1), family tutoring conducted on an additional day per week is associated with 0.053 standard deviations decreases of students' Chinese test score rank. However, this negative association is only 0.029 standard deviations in column (3) after controlling lagged rank of Chinese test. This indicates that the OLS estimate is likely biased downward because students with lower prior achievement are more inclined to receive family tutoring, and once previous performance is accounted for, the estimated negative effect of tutoring becomes substantially smaller. Moreover, comparing with column (3), column (5) adds lagged family tutoring, and the negative association reduces to 0.02 standard deviation. Controlling for past tutoring absorbs the long-term reverse causality arising from earlier poor performance, which makes the contemporaneous tutoring coefficient move in a more positive direction.

Private tutoring (PT) exhibits a different pattern. In the contemporaneous specification, PT is negatively correlated with all three subjects, but the effect either disappears or becomes positive in the alternative models. For example, in Panel B column (2), one more hour of private tutoring per week has a positive and statistically significant effect of 0.006 standard deviations on a student's mathematics test rank, whereas the lagged PT coefficient is negative in both the CU and CVA models. This suggests that private tutoring may improve exam rankings in math once unobserved student heterogeneity is controlled for, but the negative lagged associations again point to reverse causality, where tutoring is introduced after weak performance.

Taken together, the results for school tests imply that family tutoring offers little benefit, while private tutoring can raise relative exam rankings in certain subjects, especially mathematics, although this effect is sensitive to model choice.

### 1.5.2 Cognitive ability VS Test-taking ability

Since the effects on school test outcomes have been decomposed into cognitive ability and test-taking ability in Section 1.3.3, the regressions in Table 1.3 examine whether tutoring influences students' test outcomes through development in cognitive ability or through the improvements of test-taking ability. We expect both cognitive ability and test-taking ability to be less affected by negative selection than school test outcomes, because parents can directly observe only their children's school test scores or ranks, but not their cognitive ability rank or test-taking ability rank.

As shown in Table 1.3 panel A, the results for standardized cognitive test outcomes are more uniformly negative. Both family tutoring and private tutoring are associated with lower cognitive scores across contemporaneous and lagged specifications. For example, in column (3), family tutoring conducted on an additional day per week reduces cognitive test rank by about 0.033 in the VA model, while one more weekly hour of private tutoring reduces scores by 0.007 standard deviation in the cumulative model. These negative associations weaken in the CVA model, where lagged outcomes are included, suggesting that part of the pattern reflects reverse causality and omitted prior achievement.

However, in panel B, the estimates show that private tutoring (in current wave) is positively and significantly associated with test-taking ability in most models. For example, one more hour of private tutoring per week positively increase students' test taking ability around 0.007 to 0.005 standard deviations in FE and CU model. In the VA, and CVA model, one additional hour of private tutoring per week raises test-taking ability by about 0.006-0.008 standard deviations. However, family tutoring has no significant effect on students' test-taking ability.

Taken together, these findings indicate that private tutoring enhances test-taking ability that improve students' relative rank in school assessments, whereas parental involvement has no positive effect on students' test-taking ability, as parents generally possess limited knowledge of school-specific exam requirements. However, neither private nor family tutoring can promote cognitive development, and in the long run, such practices may even impair students' creativity (Han & Suh, 2023). This negative effect can also be driven by reduced sleep and difficulties in maintaining concentration (Könen et al., 2015; Zhang, 2023).

**Table 1.2:** Effects of private tutoring and family tutoring on students' school test outcomes.

| | (1) Contemporaneous | (2) FE | (3) VA | (4) CU | (5) CVA |
|---|---|---|---|---|---|
| **Panel A: Chinese test score rank** | | | | | |
| Private tutoring | -0.008*** | 0.000 | 0.000 | -0.001 | 0.001 |
| | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.009*** | -0.003 |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.053*** | 0.001 | -0.029*** | -0.044*** | -0.020** |
| | (0.009) | (0.007) | (0.009) | (0.014) | (0.010) |
| L1.Family tutoring | | | | -0.044*** | -0.021** |
| | | | | (0.012) | (0.008) |
| **Panel B: Math test score rank** | | | | | |
| Private tutoring | -0.005** | 0.006*** | 0.005** | 0.002 | 0.006*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.011*** | -0.006*** |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.055*** | -0.004 | -0.019** | -0.038*** | -0.015 |
| | (0.009) | (0.008) | (0.009) | (0.012) | (0.010) |
| L1.Family tutoring | | | | -0.041*** | -0.008 |
| | | | | (0.014) | (0.009) |
| **Panel C: English test score rank** | | | | | |
| Private tutoring | -0.005** | 0.002 | 0.001 | 0.002 | 0.002 |
| | (0.002) | (0.001) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.010*** | -0.004** |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.051*** | 0.003 | -0.022*** | -0.053*** | -0.016* |
| | (0.010) | (0.007) | (0.008) | (0.014) | (0.008) |
| L1.Family tutoring | | | | -0.036** | -0.015* |
| | | | | (0.015) | (0.009) |
| **Panel D: Chinese plus math test score rank** | | | | | |
| Private tutoring | -0.007*** | 0.003** | 0.004** | 0.000 | 0.005*** |
| | (0.002) | (0.001) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.011*** | -0.003* |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.061*** | -0.005 | -0.020** | -0.045*** | -0.017* |
| | (0.009) | (0.007) | (0.008) | (0.014) | (0.009) |
| L1.Family tutoring | | | | -0.046*** | -0.009 |
| | | | | (0.014) | (0.008) |
| Control | YES | YES | YES | YES | YES |
| Student FE | NO | YES | NO | NO | NO |
| Wave FE | NO | YES | NO | NO | NO |
| N | 6,838 | 6,838 | 3,419 | 3,419 | 3,419 |
| Number of students | | 3,419 | | | |

**Note:** Data from CEPS 2013-2015. Controls include hukou type, healthy status, whether living with father and mother, whether boarding at school, number of siblings, father and mother's occupation and education, family economic status, Chinese, Math, English teachers' education level and teaching experience, and class size. Standard errors in parentheses, and clustered by school. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

**Table 1.3:** Effects of private tutoring and family tutoring on students' cognitive ability and test-taking ability.

| | (1) Contemporaneous | (2) FE | (3) VA | (4) CU | (5) CVA |
|---|---|---|---|---|---|
| **Panel A: Cognitive test score rank** | | | | | |
| Private tutoring | -0.010*** | -0.004** | -0.007*** | -0.007** | -0.006*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.007** | -0.003 |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.043*** | 0.002 | -0.033*** | -0.036*** | -0.023* |
| | (0.009) | (0.011) | (0.012) | (0.013) | (0.012) |
| L1.Family tutoring | | | | -0.036*** | -0.024** |
| | | | | (0.013) | (0.012) |
| **Panel B: Test taking ability rank** | | | | | |
| Private tutoring | -0.002 | 0.007*** | 0.006*** | 0.005** | 0.008*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.010*** | -0.007*** |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.046*** | -0.008 | -0.024** | -0.032** | -0.020* |
| | (0.008) | (0.009) | (0.011) | (0.014) | (0.011) |
| L1.Family tutoring | | | | -0.032** | -0.010 |
| | | | | (0.014) | (0.011) |
| Control | YES | YES | YES | YES | YES |
| Student FE | NO | YES | NO | NO | NO |
| Wave FE | NO | YES | NO | NO | NO |
| N | 6,838 | 6,838 | 3,419 | 3,419 | 3,419 |
| Number of students | | 3,419 | | | |

**Note:** Data from CEPS 2013-2015. Controls include hukou type, healthy status, whether living with father and mother, whether boarding at school, number of siblings, father and mother's occupation and education, family economic status, Chinese, Math, English teachers' education level and teaching experience, and class size. Standard errors in parentheses, and clustered by school. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

## 1.6 Robustness check

In this section, we examine whether our results are sensitive to the scaling of outcome variables or to sample selection. Table 1.4 reports the results obtained when replacing the normalized rank with a uniformly distributed rank. Table 1.5 and 1.6 present results based on alternative samples that include students admitted through school selection and class assignment, respectively.

First, we examine how the choice of normalisation affects the results. Since ranks are often expressed from 1 to N, or in percentile form in practice, we re-scale the rank variable in Table

1.4. After replacing the normalized rank with the uniformly distributed rank, most coefficients become smaller, although the statistical significance of most estimates remains unchanged. The uniformly distributed rank may magnify the effects of tutoring for students positioned around the middle of the distribution in terms of their academic or cognitive/test-taking ability ranks. This may occur because the normalized rank is more compressed in the middle range, so an equivalent change in position represents a smaller difference than in the uniform distribution. However, this pattern is not observable in Table 1.4, as the uniformly distributed rank, defined on the percentile scale $(0, 1)$, has a much smaller standard deviation than the normalized rank. The standard deviation of the uniformly distributed rank is 0.288, substantially smaller than 1, which mechanically leads to smaller coefficient estimates compared with those outcomes in Table 1.2 and Table 1.3.

Second, we check the decision of sample selection, and examine whether including students whose parents selected their schools alters the results. For these students, parents typically choose higher-quality schools, and they are therefore positively selected in the sense that they are likely to receive more educational inputs and to perform better. However, Table 1.5 shows that most coefficients remain similar after including these positively selected students. In the fixed-effects specification, one extra hour of private tutoring even produces a smaller benefit for these students (0.05 standard deviations in Table 1.5 compared with 0.07 in Table 1.3). This pattern arises because enrolling in a better school also implies holding a relatively lower rank within that school. Additional inputs, such as tutoring or other educational support, do not yield advantages relative to their school peers, and the efficiency of tutoring declines when inputs become excessive. As a result, the advantage from positive selection appears only sufficient to offset differences in learning efficiency between these students and their classmates.

Last, we examine an additional sample selection decision in Table 1.6 by including schools that do not randomly assign students to classes. In these schools, the average ability of students may differ across classes. Parents with children in low or regular classes may adjust their educational inputs in different ways. If they care about their child's rank within the class, they may reduce investment because their child performs relatively well. If they focus instead on the systematic gap between their child and pupils in high classes, they may increase tutoring investment, which could increase negative selection. In Table 1.6, when these schools are included, we do not observe substantial changes in most coefficients. This suggests that parents with children in regular classes are not systematically increasing or decreasing their tutoring inputs in a way that meaningfully affects the estimates.

**Table 1.4:** Robustness check: uniform-distributed rank.

| **Uniform-distributed rank** | | | | | |
|---|---|---|---|---|---|
| | (1) Contemporaneous | (2) FE | (3) VA | (4) CU | (5) CVA |
| **Panel A: Chinese test score rank** | | | | | |
| Private tutoring | -0.002*** | -0.000 | 0.000 | -0.000 | 0.000 |
| | (0.001) | (0.000) | (0.001) | (0.001) | (0.001) |
| L1.Private tutoring | | | | -0.003*** | -0.001 |
| | | | | (0.001) | (0.001) |
| Family tutoring | -0.016*** | 0.000 | -0.009*** | -0.013*** | -0.006** |
| | (0.002) | (0.002) | (0.003) | (0.004) | (0.003) |
| L1.Family tutoring | | | | -0.013*** | -0.006** |
| | | | | (0.004) | (0.002) |
| **Panel B: Math test score rank** | | | | | |
| Private tutoring | -0.002** | 0.002*** | 0.001** | 0.001 | 0.002*** |
| | (0.001) | (0.000) | (0.001) | (0.001) | (0.001) |
| L1.Private tutoring | | | | -0.003*** | -0.002*** |
| | | | | (0.001) | (0.001) |
| Family tutoring | -0.017*** | -0.002 | -0.005* | -0.010*** | -0.004 |
| | (0.003) | (0.002) | (0.003) | (0.004) | (0.003) |
| L1.Family tutoring | | | | -0.013*** | -0.004 |
| | | | | (0.004) | (0.003) |
| **Panel C: English test score rank** | | | | | |
| Private tutoring | -0.001** | 0.001 | 0.000 | 0.001 | 0.001 |
| | (0.001) | (0.000) | (0.001) | (0.001) | (0.001) |
| L1.Private tutoring | | | | -0.003*** | -0.001** |
| | | | | (0.001) | (0.000) |
| Family tutoring | -0.015*** | 0.001 | -0.007*** | -0.016*** | -0.005* |
| | (0.003) | (0.002) | (0.002) | (0.004) | (0.002) |
| L1.Family tutoring | | | | -0.011*** | -0.005** |
| | | | | (0.004) | (0.002) |
| **Panel D: Cognitive test score rank** | | | | | |
| Private tutoring | -0.003*** | -0.001** | -0.002*** | -0.002** | -0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| L1.Private tutoring | | | | -0.002*** | -0.001 |
| | | | | (0.001) | (0.001) |
| Family tutoring | -0.013*** | -0.001 | -0.009** | -0.010*** | -0.006* |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) |
| L1.Family tutoring | | | | -0.010** | -0.006* |
| | | | | (0.004) | (0.003) |
| **Panel E: Test taking ability rank** | | | | | |
| Private tutoring | -0.001* | 0.002*** | 0.002** | 0.001* | 0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| L1.Private tutoring | | | | -0.003*** | -0.002*** |
| | | | | (0.001) | (0.001) |
| Family tutoring | -0.014*** | -0.001 | -0.008** | -0.009** | -0.006* |
| | (0.002) | (0.003) | (0.003) | (0.004) | (0.003) |
| L1.Family tutoring | | | | -0.011*** | -0.005 |
| | | | | (0.004) | (0.003) |
| Control | YES | YES | YES | YES | YES |
| Student FE | NO | YES | NO | NO | NO |
| Wave FE | NO | YES | NO | NO | NO |
| N | 6,838 | 6,838 | 3,419 | 3,419 | 3,419 |
| Number of students | | 3,419 | | | |

**Note:** Data from CEPS 2013-2015. Controls include hukou type, healthy status, whether living with father and mother, whether boarding at school, number of siblings, father and mother's occupation and education, family economic status, Chinese, Math, English teachers' education level and teaching experience, and class size. Standard errors in parentheses, and clustered by school. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

**Table 1.5:** Robustness check: including school selection samples.

| **Including school selection** | | | | | |
|---|---|---|---|---|---|
| | (1) Contemporaneous | (2) FE | (3) VA | (4) CU | (5) CVA |
| **Panel A: Chinese test score rank** | | | | | |
| Private tutoring | -0.007*** | -0.001 | -0.001 | -0.003 | -0.000 |
| | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) |
| L1.Private tutoring | | | | -0.006*** | -0.000 |
| | | | | (0.002) | (0.001) |
| Family tutoring | -0.052*** | -0.004 | -0.029*** | -0.047*** | -0.023*** |
| | (0.008) | (0.006) | (0.008) | (0.013) | (0.008) |
| L1.Family tutoring | | | | -0.035*** | -0.014* |
| | | | | (0.011) | (0.007) |
| **Panel B: Math test score rank** | | | | | |
| Private tutoring | -0.006*** | 0.004*** | 0.003 | -0.001 | 0.004** |
| | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.008*** | -0.004** |
| | | | | (0.003) | (0.002) |
| Family tutoring | -0.060*** | -0.004 | -0.022*** | -0.037*** | -0.016* |
| | (0.008) | (0.007) | (0.008) | (0.012) | (0.009) |
| L1.Family tutoring | | | | -0.049*** | -0.013* |
| | | | | (0.012) | (0.008) |
| **Panel C: English test score rank** | | | | | |
| Private tutoring | -0.005*** | 0.001 | 0.001 | -0.002 | 0.001 |
| | (0.002) | (0.001) | (0.001) | (0.002) | (0.001) |
| L1.Private tutoring | | | | -0.006** | -0.002 |
| | | | | (0.002) | (0.001) |
| Family tutoring | -0.051*** | -0.002 | -0.021*** | -0.045*** | -0.016** |
| | (0.008) | (0.006) | (0.006) | (0.013) | (0.007) |
| L1.Family tutoring | | | | -0.037*** | -0.011 |
| | | | | (0.013) | (0.008) |
| **Panel D: Cognitive test score rank** | | | | | |
| Private tutoring | -0.010*** | -0.004** | -0.009*** | -0.009*** | -0.008*** |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.007*** | -0.004** |
| | | | | (0.002) | (0.002) |
| Family tutoring | -0.046*** | -0.008 | -0.046*** | -0.047*** | -0.037*** |
| | (0.008) | (0.010) | (0.010) | (0.011) | (0.010) |
| L1.Family tutoring | | | | -0.032*** | -0.021** |
| | | | | (0.011) | (0.010) |
| **Panel E: Test taking ability rank** | | | | | |
| Private tutoring | -0.003* | 0.005*** | 0.005*** | 0.003 | 0.006*** |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.005** | -0.003 |
| | | | | (0.002) | (0.002) |
| Family tutoring | -0.046*** | -0.003 | -0.019* | -0.026* | -0.013 |
| | (0.008) | (0.008) | (0.010) | (0.013) | (0.011) |
| L1.Family tutoring | | | | -0.037*** | -0.015 |
| | | | | (0.013) | (0.010) |
| Control | YES | YES | YES | YES | YES |
| Student FE | NO | YES | NO | NO | NO |
| Wave FE | NO | YES | NO | NO | NO |
| N | 9,466 | 9,466 | 4,733 | 4,733 | 4,733 |
| Number of students | | 4,733 | | | |

**Note:** Data from CEPS 2013-2015. Controls include hukou type, healthy status, whether living with father and mother, whether boarding at school, number of siblings, father and mother's occupation and education, family economic status, Chinese, Math, English teachers' education level and teaching experience, and class size. Standard errors in parentheses, and clustered by school. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

**Table 1.6:** Robustness check: including class assignment samples.

| **Including class assignment** | | | | | |
|---|---|---|---|---|---|
| | (1) Contemporaneous | (2) FE | (3) VA | (4) CU | (5) CVA |
| **Panel A: Chinese test score rank** | | | | | |
| Private tutoring | -0.007*** | -0.000 | -0.001 | -0.001 | 0.000 |
| | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.009*** | -0.003* |
| | | | | (0.002) | (0.001) |
| Family tutoring | -0.050*** | 0.006 | -0.023*** | -0.038*** | -0.014 |
| | (0.008) | (0.007) | (0.008) | (0.013) | (0.009) |
| L1.Family tutoring | | | | -0.043*** | -0.022*** |
| | | | | (0.011) | (0.008) |
| **Panel B: Math test score rank** | | | | | |
| Private tutoring | -0.006*** | 0.004*** | 0.003* | 0.001 | 0.005** |
| | (0.002) | (0.001) | (0.002) | (0.003) | (0.002) |
| L1.Private tutoring | | | | -0.010*** | -0.005*** |
| | | | | (0.002) | (0.002) |
| Family tutoring | -0.049*** | 0.004 | -0.009 | -0.025** | -0.003 |
| | (0.008) | (0.007) | (0.009) | (0.012) | (0.009) |
| L1.Family tutoring | | | | -0.046*** | -0.014* |
| | | | | (0.012) | (0.009) |
| **Panel C: English test score rank** | | | | | |
| Private tutoring | -0.004** | 0.002* | 0.001 | 0.002 | 0.002 |
| | (0.002) | (0.001) | (0.001) | (0.003) | (0.001) |
| L1.Private tutoring | | | | -0.009*** | -0.004*** |
| | | | | (0.002) | (0.001) |
| Family tutoring | -0.045*** | 0.005 | -0.018** | -0.044*** | -0.011 |
| | (0.009) | (0.007) | (0.008) | (0.013) | (0.009) |
| L1.Family tutoring | | | | -0.036*** | -0.015* |
| | | | | (0.013) | (0.008) |
| **Panel D: Cognitive test score rank** | | | | | |
| Private tutoring | -0.008*** | -0.002 | -0.007*** | -0.006** | -0.005*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.009*** | -0.005*** |
| | | | | (0.002) | (0.002) |
| Family tutoring | -0.037*** | 0.004 | -0.033*** | -0.032*** | -0.022** |
| | (0.008) | (0.010) | (0.010) | (0.011) | (0.010) |
| L1.Family tutoring | | | | -0.036*** | -0.026** |
| | | | | (0.013) | (0.011) |
| **Panel E: Test taking ability rank** | | | | | |
| Private tutoring | -0.003* | 0.004** | 0.005** | 0.004* | 0.006*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| L1.Private tutoring | | | | -0.008*** | -0.004** |
| | | | | (0.002) | (0.002) |
| Family tutoring | -0.043*** | 0.002 | -0.012 | -0.020 | -0.005 |
| | (0.008) | (0.009) | (0.010) | (0.013) | (0.011) |
| L1.Family tutoring | | | | -0.039*** | -0.017* |
| | | | | (0.012) | (0.010) |
| Control | YES | YES | YES | YES | YES |
| Student FE | NO | YES | NO | NO | NO |
| Wave FE | NO | YES | NO | NO | NO |
| N | 8,532 | 8,532 | 4,266 | 4,266 | 4,266 |
| Number of students | | 4,266 | | | |

**Note:** Data from CEPS 2013-2015. Controls include hukou type, healthy status, whether living with father and mother, whether boarding at school, number of siblings, father and mother's occupation and education, family economic status, Chinese, Math, English teachers' education level and teaching experience, and class size. Standard errors in parentheses, and clustered by school. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

## 1.7 Discussion

This study shows a clear difference between family and private tutoring in China's exam-oriented system. Family tutoring has limited or even negative effects on both school tests and cognitive outcomes, consistent with its role as a remedial rather than productive input. Private tutoring, by contrast, improves students' school test ranks mainly by enhancing test-taking ability, but it does not improve cognitive skills.

These results suggest that private tutoring improves students' performance in school examinations primarily by enhancing test-taking ability rather than cognitive skills. Because cognitive ability is more closely related to deeper learning and long-run skill formation, this pattern implies that tutoring may improve short-term exam performance without necessarily strengthening the underlying skills that support long-run educational development. Policy interventions should therefore focus on reducing excessive reliance on private tutoring and strengthening the capacity of schools to train both cognitive skills and test-taking skills.

Three limitations remain. First, reverse causality cannot be fully excluded, as parents may increase tutoring after poor performance. Second, without individual fixed-effect, the test-taking ability may include biases caused by the interaction of household/personal characteristics and the cognitive ability. Third, the two-wave CEPS panel restricts our ability to trace long-run effects. Future work using richer data is needed to better establish causal impacts.

# References

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? accounting for learning dynamics. *American Economic Journal: Applied Economics*, *3*(3), 29–54.

Baker, D. P., Akiba, M., LeTendre, G. K., & Wiseman, A. W. (2001). Worldwide shadow education: Outside-school learning, institutional quality of schooling, and cross-national mathematics achievement. *Educational Evaluation and Policy Analysis*, *23*(1), 1–17.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9–49.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, *75*(299), 493–517.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354–13359.

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, american style: Test preparation, the sat and college enrollment. *Social Forces*, *89*(2), 435–461.

Byun, S. Y. (2014). Shadow education and academic success in republic of korea. In *Korean education in changing economic and demographic contexts* (pp. 39–58). Springer.

Carneiro, P., & Rodrigues, M. (2009). *Evaluating the effect of maternal time on child development using the generalized propensity score* (tech. rep. No. 12th IZA European Summer School in Labor Economics). Institute for the Study of Labor.

Dang, H.-A. (2007). The determinants and impact of private tutoring classes in vietnam. *Economics of Education Review*, *26*(6), 683–698.

Datcher-Loury, L. (1988). Effects of mother's home time on children's schooling. *The Review of Economics and Statistics*, 367–373.

Del Boca, D., Monfardini, C., & Nicoletti, C. (2012). *Children's and parent's time-use choices and cognitive development during adolescence* (tech. rep. No. 6). Human Capital and Economic Opportunity Working Group.

Del Bono, E., Francesconi, M., Kelly, Y., & Sacker, A. (2016). Early maternal time investment and early child outcomes. *The Economic Journal*, *126*(596), F96–F135.

Fiorini, M., & Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, *32*(4), 787–836.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308.

Guo, Y., Chen, Q., Zhai, S., & Pei, C. (2020). Does private tutoring improve student learning in china? evidence from the china education panel survey. *Asia & the Pacific Policy Studies*, *7*(3), 322–343.

Han, S., & Suh, H. (2023). The effects of shadow education on high school students' creative thinking and academic achievement in mathematics: The case of the republic of korea. *Educational Studies*, *49*(2), 314–333.

Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of econometrics*, *121*(1-2), 39–98.

Hanushek, E. A. (2020). Education production functions. In *The economics of education* (pp. 161–170). Academic Press.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–668.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464.

Hill, C. R., & Stafford, F. P. (1974). Allocation of time to preschool children and educational opportunity. *Journal of Human Resources*, 323–341.

Hill, C. R., & Stafford, F. P. (1980). Parental care of children: Time diary estimates of quantity, predictability, and variety. *Journal of Human Resources*, 219–239.

Kang, C., & Park, Y. (2021). Private tutoring and distribution of student academic outcomes: An implication of the presence of private tutoring for educational inequality. *Korean Economic Review*, *37*(2), 287–326.

Könen, T., Dirk, J., & Schmiedek, F. (2015). Cognitive benefits of last night's sleep: Daily variations in children's sleep behavior are related to working memory fluctuations. *Journal of Child Psychology and Psychiatry*, *56*(2), 171–182.

Liao, X., & Huang, X. (2018). Who is more likely to participate in private tutoring and does it work?: Evidence from pisa (2015). *ECNU Review of Education*, *1*(3), 69–95.

Loyalka, P., & Zakharov, A. (2016). Does shadow education help students prepare for college? evidence from russia. *International Journal of Educational Development*, *49*, 22–30.

Reyes, G. (2023). Cognitive endurance, talent selection, and the labor market returns to human capital. *arXiv preprint arXiv:2301.02575*.

Soloman, S. R., & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods*, *8*(2), 9.

Sun, L., Shafiq, M. N., McClure, M., & Guo, S. (2020). Are there educational and psychological benefits from private supplementary tutoring in mainland china? evidence from the china education panel survey, 2013–15. *International Journal of Educational Development*, *72*, 102144.

Thomsen, M. K. (2015). Parental time investments in children: Evidence from denmark. *Acta Sociologica*, *58*(3), 249–263.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, *1*(1), 91–136.

Villena-Roldan, B., & Ríos-Aguilar, C. (2012). *Causal effects of maternal time-investment on children's cognitive outcomes* (tech. rep. No. 285). Center for Applied Economics.

Yang, J., & Zhao, X. (2020). Parenting styles and children's academic performance: Evidence from middle schools in china. *Children and Youth Services Review*, *113*, 105017.

Zhang, Y. (2011). *The determinants of national college entrance exam performance in china-with an analysis of private tutoring* [Doctoral dissertation, Columbia University].

Zhang, Y. (2013). Does private tutoring improve students' national college entrance exam performance?—a case study from jinan, china. *Economics of Education Review*, *32*, 1–28.

Zhang, Y. (2023). Time spent on private tutoring and sleep patterns of chinese adolescents: Evidence from a national panel survey. *Children*, *10*(7), 1231.

Zhang, Y., & Liu, J. (2016). The effectiveness of private tutoring in china with a focus on class-size. *International Journal of Educational Development*, *46*, 35–42.

Zhong, J., He, Y., Gao, J., Wang, T., & Luo, R. (2020). Parenting knowledge, parental investments, and early childhood development in rural households in western china. *International Journal of Environmental Research and Public Health*, *17*(8), 2792.

# Chapter 2

# Effect of school meal policy on parents' outcomes: evidence from China

**Ziyi Huang*** **Qingxu Yang**†

**Abstract:** *This paper studies the effect of the availability of school meals and meal subsidies on parents' leisure time allocation and labour market participation in China. We use the Chinese General Social Survey (CGSS) and China Household Finance Survey (CHFS) to implement a Difference-in-Difference imputation method to test the effect of the roll-out of school meal availability and school meal subsidy in China. Our research shows that the policy has spillover effects on parents who benefit from released time constraints in several ways. Specifically, we show that more availability of school meal increases urban* hukou *fathers' time spent listening music and meals with subsidy increases rural* hukou *parents' time spent watching TV. Labour supply responses differ by education level, with low-education gaining increase their working hours, while highly educated urban* hukou *parents have no significant effects. These results suggest that providing school meals narrows socioeconomic gaps in parents' welfare.*

**Keywords:** School meal, Time Allocation and labour Supply, Policy analysis, Government welfare, China

**JEL code:** H53, I20, I28, I38, J22

*University of Essex. Corresponding author: Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ. zh17565@essex.ac.uk

†Peking University.

## 2.1 Introduction

School meal programs have been acknowledged as an effective policy tool for strengthening human capital accumulation by enhancing children's nutritional intake during critical periods of physical growth and cognitive development. In the context of China, existing empirical literature has primarily examined effects of these programs on students' health outcomes, cognitive performance, and household educational investment (Fang & Zhu, 2022; Liang et al., 2022; H. Wang & Cheng, 2022; J. Wang et al., 2022). These studies provide compelling evidence that improved nutrition contributes to higher academic achievement and plays a role in reducing educational disparities. In addition to these direct impacts, our paper emphasizes the indirect welfare implications of school meal provision—specifically, its role as an in-kind transfer that relaxes household time constraints. This effect is particularly significant for low-income households, where the substitution of public for private expenditure on child nutrition generates non-negligible improvements in household welfare.

China's school meal policy has evolved through a combination of local and central government's efforts, with significant institutional changes occurring over time. Before 2011, provision was mainly organised by local governments and showed substantial variation. Local programmes differed across counties in nutritional standards and subsidy levels. They also changed over time, and free or subsidised meals were offered only on a limited scale, mainly for rural boarding students. In 2011, the central government launched the Student Nutrition Improvement Programme (SNIP). This marked a shift towards a more standardised and centrally coordinated system aimed at improving nutrition and promoting educational equity. The SNIP first targeted 699 counties with low economic development and expanded to more than 1,600 counties by 2018.

Under the SNIP framework, students in participating counties who hold rural *hukou* [3], including those residing in urban areas, are eligible for fully or heavily subsidized school meals. In our analysis, we categorize these students as the "meal + subsidy" treatment group. In contrast, students with urban *hukou* enrolled in participating counties may access the same school

[3] In China, "hukou" is a type of identity authentication.

meal services without financial subsidy. We describe these as the "meal only" treatment group. This institutional variation allows us to identify heterogeneous treatment effects by distinguishing between two types of policy benefits: subsidized meal provision ("meal + subsidy") and non-subsidized provision ("meal only"). Leveraging this policy design, we estimate difference-in-differences (DID) models that exploit the staggered rollout of both general meal availability and subsidized meal availability across counties over time. Given that standard two-way fixed effects estimators may produce biased estimates in the presence of treatment effect heterogeneity across time and counties, we adopt the imputation method developed by Borusyak et al. (2024), which provides inference robust to this staggered adoption setting.

We apply the DID imputation method using two nationally representative datasets: the Chinese General Social Survey (CGSS) and the China Household Finance Survey (CHFS). The CGSS is a repeated cross-sectional survey with samples clustered at the county level, covering the years 2010, 2011, 2012, 2013, 2015, and 2021. In contrast, the CHFS is a biennial panel dataset, providing personal-level longitudinal data from 2011 to 2019. For the empirical analysis, we draw on CGSS to examine changes in parental time allocation in leisure activities, while CHFS is used to estimate effects on labour market participation. Furthermore, we explore heterogeneous treatment effects by disaggregating the sample by gender and educational attainment of the parents.

In many Chinese households, preparing daily meals requires a substantial amount of time. Wei (2019) show that more than 75% of households cook at least twice per day, with each cooking episode typically lasting around 30 minutes, implying roughly one hour per day spent on cooking activities for the majority of households. This suggests that school meal provision has the potential to reduce household production time by several hours per week, thereby relaxing time constraints and allowing for a meaningful increase in labour supply. Consistent with this mechanism, our empirical results indicate that increased availability of school meals for children leads to significant changes in parental behaviour across both urban and rural households. Specifically, we find that availability of school meals reduces the time parents spend on household production such as shopping, which in turn increases their leisure time and labour supply. Among urban *hukou* parents, the policy induces a rise in working hours, accompanied by greater time spent on leisure activities such as listening to music. For rural *hukou* parents, receipt of meals with subsidy not only leads to longer working hours but also correlates with a higher frequency of television viewing and listening to music. These findings underscore the presence of meaningful spillover effects from school meal provision to the wider household, suggesting that such policies generate welfare gains beyond the direct beneficiaries. In particular, the increase in parental leisure time can be considered as the social welfare, especially in a context like China, where average working hours exceed those of all OECD countries. [4]

This study contributes to the literature by investigating the spillover effects of school meal policies on parental time allocation, encompassing both labour market participation and leisure. While a small but growing number of studies have begun to examine the implications of such policies for parental employment, working hours, and income (Fang & Zhu, 2022; G. Wang et al., 2024), most do not account for parental leisure as part of their analysis. Departing from the predominant focus on direct benefits to students, our analysis highlights how the Student Nutrition Improvement Program (SNIP) indirectly alleviates parental time, thereby enhancing household welfare. By exploiting variation in meal availability and subsidy eligibility across *hukou* categories, we identify heterogeneous treatment effects and offer new insights into the role of policy intensity. Our findings underscore the broader socioeconomic returns of child-focused implementations and call for a more comprehensive approach to policy evaluation that includes both direct and indirect welfare effects.

The rest of the paper proceeds as follows. Section 2 discusses the related literature. Section 3 is the institutional background. Section 4 is the definition of the treatment. Section 5 provides a framework. Section 6 describes the data and provides descriptive statistics. Section 7 presents the estimation strategy and regressions for the event study. Section 8 shows the main results. Section 9 is robustness check. Section 10 offers heterogeneity, and section 11 is the conclusion.

[4] In 2022, Chinese average weekly working hour is 47.9 hours. Source from Chinese government, available at https://www.gov.cn/xinwen/2023-01/17/content_5737627.htm?eqid=d54ae40200027f700000000264672041

## 2.2 Literature

### 2.2.1 Types of school meal treatment effects

In the literature on school meal policies, policy implementations are typically analysed in terms of how they alter access to free school meals. Three primary types of changes in access are commonly examined.

First, the implementation of the school meal policy can result in individuals transitioning from receiving nothing to receiving means-tested meals. A prominent example is the School Breakfast Program (SBP) in the United States, a federally funded entitlement program that offers breakfast to all students in participating schools. Under this scheme, students from low-income households are eligible for free or heavily subsidized meals, whereas others may be required to pay a full price. Related studies exploit this variation in eligibility to identify causal effects: Frisvold (2015) uses income thresholds that vary across states as a source of identification, while Bhattacharya et al. (2006) leverage the difference between school and non-school periods to assess the program's impact.

Second, some other literature focuses on policies shifting from receiving means-tested to receiving universal school meals. Notable examples include the Universal Infant Free School Meals (UIFSM) policy in the UK (Holford & Rabe, 2022, 2024), the Community Eligibility Provision (CEP) in the US (Marcus & Yewell, 2022; Rothbart et al., 2022), and various Universal Free Meals (UFM) initiatives in the US (Schwartz & Rothbart, 2020). In these settings, low-income students already had access to free meals prior to the policy change, and the reform extends eligibility to all students, thereby removing the income-based eligibility requirement.

Third, some policies represent a transition from receiving nothing to receiving universal school meals. Examples include the School Meals Programme (SMP) in Ethiopia (Poppe et al., 2019), school feeding initiatives in India (Chakraborty & Jayaraman, 2019), and the introduction of universal school meals in Sweden in the 1960s (Lundborg et al., 2022). These policies introduce free school meals for all students in settings where no such provision previously existed.

Finally, a smaller body of literature examines reforms that do not alter access but change the quality or nutritional composition of meals. Examples include the Feed Me Better campaign in the UK (Belot & James, 2011) and pre-WWII school meal reforms in Norway (Bütikofer et al., 2018). In these cases, eligibility for free school meals remains unchanged, but the content of the meals is modified following the implementation.

In China, a growing body of literature has examined the Student Nutrition Improvement Program (SNIP), with most studies only focusing on students in rural area (Fang & Zhu, 2022; Liang et al., 2022; G. Wang et al., 2024; H. Wang & Cheng, 2022; J. Wang et al., 2022; Zheng et al., 2022). These studies typically characterize the SNIP as a transition from receiving nothing to receiving universal school meals for students in rural China, which align with the framework commonly used in the international literature.

In our study, the policy is not geographically confined to rural areas[5] but is instead targeted based on students' hukou identity[6]. In many counties, the implementation of the SNIP provides meal subsidies specifically to students with rural *hukou*, while also funding the infrastructure of school canteens regardless of whether the schools are located in rural or urban areas.

Accordingly, we distinguish between two treatment groups based on *hukou* status. For rural *hukou* students, SNIP is classified as from receiving nothing to receiving universal school meals implementation, combining both the availability of school meals and direct subsidies (meal + subsidy). This setting is consistent with prior literature. In contrast, for urban *hukou* students, we define the treatment as a transition from receiving nothing to being available to school meals but need to pay. i.e., meal only. This distinction is relatively underexplored in the existing literature. While urban *hukou* students do not benefit from the in-kind transfers, they may still benefit from the program, such as their parents reduced time spent on meal preparation due to the availability of school canteens.

[5]Restricting sample in rural area is an easier way to study the SNIP, since almost all students in rural area are rural *hukou*. These literature only need to consider about effects from meal subsidies, and no need to investigate complex situations (mixed of rural *hukou* and urban *hukou*) in the city.

[6]For example, in Chinese most urbanized city, Shanghai provides meal subsidy for students with rural hukou, and least urbanized areas such as Taijiang and Zhenning also provides meal subsidy for students with rural hukou even they are in the county centre (urban area). Resource from Shanghai, Taijiang, Zhenning government, available at: https://edu.sh.gov.cn/zcjd_area_3108/20200706/0015-xw_67609.html; https://www.gztaijiang.gov.cn/zwgk/zdlyxx/jy/202310/t20231024_82835601.html; https://www.gzzn.gov.cn/gk/zfgb/201705/P020241009617049385613.pdf

### 2.2.2 Effects on students and other family members

School meal programs are primarily designed to improve children's nutrition and physical health[7], but a growing body of literature has shown their broader impacts, including on educational outcomes and household behaviour. This section reviews relevant studies across five key areas: physical health, cognitive and academic performance, non-cognitive development, long-term outcomes, and household spillover effects.

Empirical evidence suggests school meals significantly improve students' physical health and development. In England, Holford and Rabe (2022, 2024) find that the UIFSM policy helps children maintain healthy body weight. In the U.S., Rothbart et al. (2022) show that CEP reduces obesity among secondary students. In China, the SNIP improves children's height, reduces illness, and boosts self-reported confidence, especially in rural areas (Liang et al., 2022; J. Wang et al., 2022). School meals also enhance academic outcomes. Studies in the U.S. and Ethiopia find that access to free or subsidized meals improves test scores (Frisvold, 2015; Poppe et al., 2019; Schwartz & Rothbart, 2020) . Meal quality matters as well, Belot and James (2011) show that healthier meals from the "Feed Me Better" campaign in the UK led to better academic performance.

In terms of non-cognitive development, Zheng et al. (2022) find that SNIP strengthens social relationships and personality traits, particularly for disadvantaged groups. Using CEPS and an instrumental variable strategy, they show improvements in peer, teacher-student, and parent-child interactions.

Long-term studies also document persistent effects into adulthood. In Norway and Sweden, school meal exposure during childhood is linked to higher educational attainment and increased lifetime earnings (Bütikofer et al., 2018; Lundborg et al., 2022), suggesting these programs function as long-term human capital investments.

Finally, school meals affect other household members. In India, siblings benefit from improved cognitive outcomes due to intra-household resource reallocation (Chakraborty & Jayaraman, 2019). In the U.S., school breakfast access improves household diet quality (Bhattacharya

[7]All programs we list above including SBP, UNIFSM, CEP, UFM, SMP, and SNIP, they all have nutritious standards.

et al., 2006). In China, SNIP reduces food spending and increases educational investment (Fang & Zhu, 2022; H. Wang & Cheng, 2022). Moreover, it enables parents, especially mothers, to work more, reallocating time from food preparation to labour market (Fang & Zhu, 2022; Holford & Rabe, 2022; James, 2024; G. Wang et al., 2024).

This paper adds to the literature by highlighting the indirect effects of school meal programs on household members. Existing studies have documented a variety of household-level spillovers beyond students' health and academic performance, including effects on food security, educational investment, and even parental labour market participation. Our study builds on this body of work by providing a more systematic examination of behavioural responses within households, with a particular focus on how school meal provision influences parental labour supply and leisure time allocation. By documenting these indirect and heterogeneous effects, the paper builds school meal policies within a broader socio-economic context and offers new evidence on their implications for household welfare.

## 2.3 Background

### 2.3.1 Schooling in China

In China, children typically begin primary education at the age of six and enrol in secondary (middle) school at the age of twelve. Both primary and secondary education constitute the nine-year compulsory education system, during which students are exempt from tuition fees. The academic year start in September and is divided into two terms, interspersed with two vacation periods. On school days, instruction generally begins between 7:00 and 8:00 a.m. and ends between 5:00 and 6:00 p.m. Most schools allocate a midday break of approximately two hours, allowing time for lunch and rest.

As of 2012, China's compulsory education system encompassed approximately 280,000 schools serving around 145 million students across 2,852 counties. Each county is further divided into multiple school districts, ranging from a few to several hundred, with students typically assigned to schools based on residential location. In urban districts or county centers, students commonly attend school as day pupils. In contrast, students residing in rural or remote areas frequently board at schools. A significant proportion of these boarding students come from economically disadvantaged households. For these students, boarding fees are waived, and they are eligible to receive additional subsidies from government programs aimed at promoting educational equity.

### 2.3.2 Early practice of school meal in China

In the absence of formal school meal programs, students typically obtain lunch by bringing food from home (packed lunches) or purchasing meals off-campus. In medium and large cities, the city of Hangzhou initiated one of the earliest pilot programs for school meal provision in 1987, involving nine schools. [8] By the late 1990s, Beijing began promoting the provision of affordable and nutritious meals to students citywide, marking a significant step toward institutionalized school meal programs.[9]

In rural and economically disadvantaged areas, charity activities—such as the "Hope Kitchen" project[10]—established school canteens in a limited number of schools prior to 2011. Similarly, before the nationwide rollout of school meal programs, few local governments proactively constructed school canteens and subsidized meals primarily for boarding students with rural *hukou*. In practice, these canteens were often accessible to non-boarding students as well. In the survey for middle schools, approximately 60% of middle schools enrol both boarding and non-boarding students,[11] and these non-boarding students are possible to be available to school meal.

### 2.3.3 The Student Nutrition Improvement Program (SNIP)

In November 2011, the Chinese central government launched the Student Nutrition Improvement Program (SNIP), a nationwide initiative aimed at enhancing the nutritional intake of students enrolled in compulsory education with rural *hukou* status. The program provided heavily subsidized meals to eligible students in designated pilot areas. By 2012, SNIP covered 699

[8] Source from Fenghuang News, available at: https://news.ifeng.com/c/7fbyXVhOozL

[9] Source from official media 'Beijing Daily', available at: https://baijiahao.baidu.com/s?id=1652043661286522182&wfr=spider&for=pc

[10] Source from government website, available at : https://www.gov.cn/jrzg/2013-05/20/content_2406772.htm

[11] Data source: China Education Panel Survey (CEPS). CEPS is a nationally representative survey with anonymized school identifier.

rural counties (out of 2,842 counties nationwide). Between 2012 and 2018, an additional 900 counties voluntarily implemented school meal programs, some of which extended or formalized earlier local initiatives.

The SNIP policy framework outlined five primary objectives. First, to initiate a national pilot program offering meal subsidies specifically for students holding rural *hukou* in underdeveloped areas—explicitly excluding students with urban *hukou*. Second, to encourage and support local governments in designing and implementing their own pilot programs, tailored to regional conditions. Third, to improve meal service infrastructure by requiring schools to construct canteens and deliver standardized meals to students with rural *hukou*, regardless of whether they are located in rural or urban areas. Fourth, to foster participation from a broad range of stakeholders, including civil society and private actors. Fifth, to increase subsidies for boarding students to further address equity concerns.

In terms of implementation, local governments serve as the primary agents of school meal provision. The central government's role is primarily fiscal: providing earmarked subsidies to participating counties. The initial subsidy was set at 3 CNY [12] per eligible rural student per day and was increased to 4 CNY in 2014. Local governments, using either central transfers or their own funds, are responsible for constructing school canteens or coordinating centralized purchase of school meals. Under this system, students with rural *hukou* are eligible to receive free or low-cost meals, while other students will be charged the full cost.

The national nutritional guidelines under the Student Nutrition Improvement Program (SNIP) are based on WS/T 554–2017,[13] although local governments are permitted to formulate region-specific nutritional standards based on local dietary habits and food availability.[14]

The Student Nutrition Improvement Program changes how students obtain their daily lunch at school. Before the implementation of the program, many students brought meals from home or returned home during the lunch break, which required households to spend time preparing food or coordinating meal arrangements during the day. Under SNIP, participating schools

[12] 3 CNY equals to 0.33 pounds, and the base year for exchange rate is 2023

[13] This standard requires school meals to provide 65g of protein and 950mg of calcium per day for a male student aged 12–14.

[14] Source from government website, available at: http://www.moe.gov.cn/srcsite/A05/s7052/202211/t20221111_984150.html

provide lunch directly to students, reducing the need for households to prepare or deliver meals. Because the policy alters how students obtain lunch, it may also affect how parents allocate time between household production, leisure, and labour market activities.

## 2.4 Definition of the treatment

In this section, we first clarify the policy used to define treatment status and describe how we control other sources of school meal provision. Next, we explain our decision to distinguish between the “Meal Only” and “Meal + Subsidy” treatments, applied respectively to students with urban and rural *hukou*. Finally, we present a figure that illustrates the coding strategy used to assign treatment status in our empirical estimations.

### 2.4.1 Policies as treatments and school canteen control

In our analysis, we define treated counties as those where school meal policies were implemented in a county-wide and universal manner, in line with the SNIP policy framework. As noted in the background section, some counties had limited forms of school meal provision prior to the national rollout of the SNIP, often through charity activities or canteens serving boarding students. Since these efforts were not uniformly implemented at the county level, they are not classified as treatment in our analysis.

To account for such historical variation, we include a control variable indicating the presence of a school canteen in a given county and year. This variable takes the value of one if any information on school canteens is found through official government sources[15], and zero otherwise. While measurement error may arise if some existing canteens are not documented in available records, the likelihood of detection increases with the actual number of canteens in a county. Therefore, counties with more school canteens are more likely to be accurately captured by this measure.

As shown in Table 2.2 and Table 2.3, more than 85 percent of the CGSS sample and over 78 percent of the CHFS sample show documented evidence of school canteens before the

[15] From government website or official media.

earliest survey year available. Given this broad historical presence of canteens, our analysis should be interpreted as capturing the expansion of school meal provision rather than its initial introduction.

### 2.4.2 Exposure to SNIP by hukou type

As previously discussed, the SNIP primarily targets students holding rural *hukou*. The government builds school canteens and provides subsidies for school meals, making these students eligible for both the availability of meals and financial support. This combined exposure for rural *hukou* students is defined as the treatment "Meal + Subsidy."

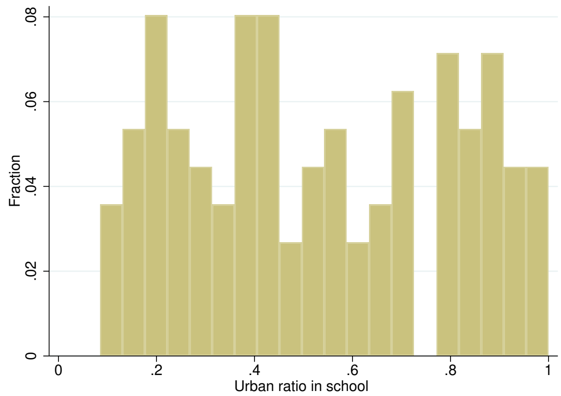
Though not subsidized under the SNIP, urban *hukou* students may still benefit from the infrastructure of the SNIP in treated counties. The reason is as follows: first, local governments provide school meals to students with rural *hukou*, regardless of whether they attend schools in rural or urban areas. As a result, schools in the treated counties that enroll students with rural *hukou* will establish school canteens. Second, we use data from the China Education Panel Survey (CEPS) to examine the proportion of students with urban *hukou* within schools, and the distribution of urban *hukou* ratio in the school is shown in Figure 2.1. CEPS, being a nationally representative survey, shows that the percentage of urban students varies significantly across schools, ranging from 10% to over 90%. In fact, only one school in the CEPS sample (less than 1%) is composed entirely of students with urban *hukou*. Based on this, we infer that nearly all urban *hukou* students in treated counties are likely to benefit from the school canteens constructed to serve students with rural *hukou*. In such cases, school meals are very likely to be available to urban *hukou* students, constituting the "Meal Only" treatment.

Additionally, around 20% of treated counties provide subsidized meals to poor students[16]. Due to data limitations, particularly the lack of individual-level income information, we do not distinguish between subsidized and non-subsidized recipients. All students in these counties are coded as "Meal Only," regardless of income eligibility. As a result, a small number of rural *hukou* students may be exposed to "Meal Only". Some of them may transition from "Meal Only" to "Meal + Subsidy" over time. In our estimation, we exclude rural *hukou* students who

[16] Poor Students are those students live in the household receiving governments' subsistence allowances. The percentage of poor household and potential measurement error will be discussed in Section 2.7.4

only receive “Meal Only” and retain only those observations where they receive the full “Meal + Subsidy” treatment.

**Figure 2.1:** The distribution of urban *hukou* ratio in the school



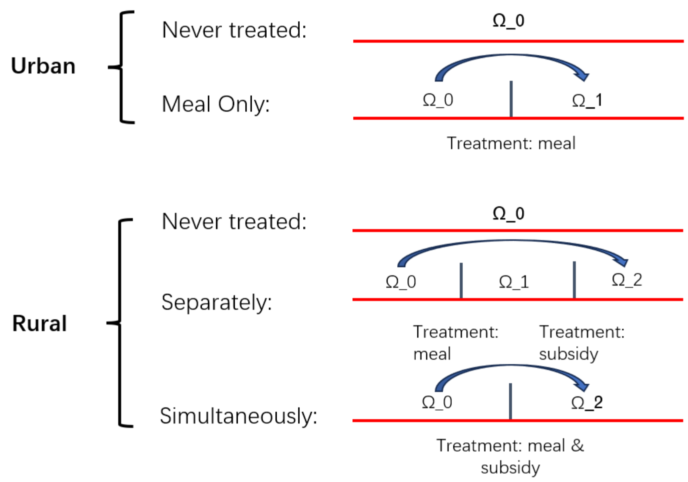

**Note:** Data from CEPS 2014. The vertical scale represents the fraction of schools, and the horizontal scale represents the ratio of urban students. Mean value of urban student ratio: 35.57%.
$Urban\,ratio\,in\,school = (number\,of\,students\,with\,urban\,hukou)/(total\,number\,of\,students)$.

### 2.4.3 Details of treatment

To the details of treatment, we are identifying county-level (county-*hukou* level) Intention-to-Treat (ITT) effects of two treatments: availability of ”Meal Only” and ”Meal + Subsidy.” Everyone potentially or partially exposed to the policy is given a treatment dummy (1 = with the policy; 0 = without the policy). For students with urban *hukou*, the treatment of ”Meal Only” is switched on if the SNIP or other local school meal policy increases their availability of receiving school meals. For students with rural *hukou*, the treatment of ”Meal + Subsidy” is switched on if the SNIP or other local school meal policy universally increases their availability to receive subsidized meals.

As shown in Figure 2.2, urban hukou students fall into two treatment categories: untreated (or pre-treatment), denoted as $\Omega_0$, and those receiving “Meal Only,” denoted as $\Omega_1$. For rural hukou students, the situation is more complex. Some counties implement school meals and subsidies sequentially (shown as “separately” in the figure), while most introduce both components simultaneously. The group receiving “Meal + Subsidy” is labeled as $\Omega_2$. For analytical consistency, we also use $\Omega_0$ as the control group for rural *hukou* students.[17]

**Figure 2.2:** The sequences of introduction of school meal policy




## 2.5 Framework

This section examines how the availability and subsidy (i.e., the effective price reduction) of school meals may influence parental allocation of time. We differentiate our analysis by *hukou* status and parental education level. Specifically, we consider two distinct groups: (1) parents with urban *hukou*, whose children receive school meals without subsidies, and (2) parents with rural *hukou*, whose children benefit from both in-school meals and direct subsidies under SNIP. Furthermore, within the urban *hukou* group, we explore heterogeneity by parental education level to assess whether responses differ according to socioeconomic status.[18] We provide hypotheses to guide our analysis.

### 2.5.1 Impact on household's lunch decision

We assume that, prior to the implementation of the school meal policy, parents make a stable choice regarding their children's lunch arrangements—either preparing lunch boxes at

[17] Ideally, we would examine the marginal effect of subsidies by analyzing counties that first implemented "Meal Only" and added subsidies later. However, the number of counties with sufficient observations in the intermediate 'Meal Only, pre-subsidy' phase is too small to permit meaningful analysis.

[18] In the next section, Table 2.2 and Table 2.3 show very few high-educated parents have rural *hukou*.

home or paying for meals at nearby restaurants. We further assume that the introduction of school meals expands the choice set but does not alter individuals' preference ordering.

With the introduction of the school meal program, the parental choice set expands from {lunchbox, restaurant} to {lunchbox, restaurant, school meal}. For urban *hukou* families, even if the school meal is priced higher than the cost of a homemade lunch, some parents may choose it due to the time-saving benefit it provides. In contrast, for rural *hukou* families,who receive both meals and subsidies under SNIP, the school meal option dominates in terms of both monetary and time savings. Therefore, we expect near-universal uptake of school meals in this group.

### 2.5.2 Impact of lunch decision change on household's constraints

The expansion of lunch options for students may relax parental constraints on time, depending on their pre-policy choices and household characteristics. For rural *hukou* parents, we assume that most could not afford to pay for restaurant meals for their children. Consequently, the introduction of subsidized school meals is expected to ease time constraints, as parents no longer need to prepare meals or purchase ingredients. In this case, we get

**Hypothesis 1:** We expect that upon receiving subsidized school meals, parents with rural *hukou* reduce the time spent on cooking and shopping, reallocating this time toward increased labour market participation and engagement in other leisure activities.

For urban *hukou* parents, the effect of school meals varies by parental education level, which we use as a proxy for income and preferences. We assume that low-educated parents have lower earnings and thus a higher marginal utility of consumption, while high-educated parents have higher earnings and a higher marginal utility of leisure. Prior to the policy, we posit that high-educated parents are more likely to allow their children to eat in restaurants, while low-educated parents tend to prepare lunch boxes at home due to cost considerations.

Among low-educated urban *hukou* parents, switching from lunch boxes to school meals primarily yields time savings. However, if school meals are more expensive than home-prepared options, they need to make a trade-off between time and money. Then we have

**Hypothesis 2:** Low-educated urban *hukou* parents exposed to the ”Meal Only” policy reduce the time spent on cooking and shopping, reallocating this time toward increased labour market participation and engagement in other leisure activities.

In contrast, for highly educated urban *hukou* parents whose children previously relied on relatively expensive restaurant meals, the programme does not ease time constraints, as these parents were not involved in meal preparation before the policy implementation.

**Hypothesis 3:** High-educated urban *hukou* parents exposed to the ”Meal Only” policy will not change their time allocation.

## 2.6 Data and sample

We combine policy implementation data on school meal programs, sourced from official government websites, with individual and household-level microdata from nationally representative social surveys in China. The linkage is based on county identifiers, allowing us to merge policy exposure with survey data at the county level. Our primary data sources are the Chinese General Social Survey (CGSS) and the China Household Finance Survey (CHFS). These datasets provide rich information on individuals’ time use, employment status, working hours, and demographic characteristics.

### 2.6.1 Data

#### Chinese General Social Survey (CGSS)

The CGSS is a nationally representative, repeated cross-sectional survey that collects individual-level data from adult respondents. The sampling is clustered at the county level, and consistent county identifiers are available across five survey waves from 2010 to 2015 (excluding 2014). In total, 134 counties appear across these five waves, of which 108 counties are observed in at least four waves and are thus included in our empirical analysis. Additionally, we incorporate data from the 2021 wave, which includes a subset of counties that overlap with those surveyed between 2010 and 2015. These counties are also retained in the regression sample to enhance temporal coverage.

**China Household Finance Survey (CHFS)**

The CHFS is a nationally representative panel dataset with a focus on intra-household information. Compared to CGSS, CHFS offers longitudinal data and more detailed information on household structure. Surveys were conducted biennially from 2011 to 2019. The initial wave in 2011 included approximately 30,000 observations, expanding to about 100,000 in later waves.[19] Consistent with the CGSS sample selection, we restrict the CHFS sample to counties that appear in at least four survey waves. A total of 257 counties meet this requirement and are included in our main regression analysis.

### 2.6.2 Observations in estimations

We restrict the dataset to counties that are repeatedly observed, and the sample is defined according to the following steps.

First, we limit the sample to individuals with children aged 6 to 15, which corresponds to the age range for compulsory education.[20][21] We assume that children in this age group are enrolled in either primary or middle school.

Second, we exclude households in which children do not live with their parents. In such cases, children are likely to attend school in a different county from that of their parents. Since the CGSS only provides geographic information for parents, this mismatch in location may lead to misclassification of treatment status and introduce identification issues. To ensure consistency between parental location and policy exposure, these households are excluded from the analysis.

[19] Approximately half of the treated counties implemented the school meal policy in 2012, which may explain the limited overlap between early waves and treated observations.

[20] We only consider whether parents have children in the compulsory education, but not the number of children. We will control children in each education level in the estimation.

[21] In CGSS, the 2021 wave only provides the number of underage children without specifying their ages. We estimate children's ages based on the parents' year of marriage.

Third, we exclude divorced parents from the sample, since their household structures and resource allocation decisions may differ substantially due to varying custodial arrangements and constraints.

Table 2.1 summarizes how school meal implementation is reflected in the survey data. The majority of school meal program adoptions occurred around 2012.[22] Accordingly, the number of observations without school meal provision declines in later waves. Since our empirical strategy relies on a difference-in-differences (DID) imputation framework, we require counties to be observed both before and after the implementation of school meals. This requirement restricts the usable treatment variation. In CHFS, although the total number of counties and observations is larger than in CGSS, many CHFS counties were not included in earlier survey waves. The second wave of CHFS (2013) was roughly three times larger than the first (2011) and coincides with the year in which many counties adopted school meals. As a result, a large share of treated counties in CHFS lack sufficient pre-treatment observations and are excluded from the sample. Consequently, despite its larger overall sample, the number of usable treatment switches in CHFS is close to CGSS.

Tables A1 through A3 in the Appendix document the cumulative rollout of school meal policies across counties in CGSS, CHFS, and the combined sample. Given that both CGSS and CHFS are nationally representative, the timing and distribution of policy exposure are broadly similar. Table 2.1 reports the final sample size by survey year after these exclusions.

[22] Table A2 in Appendix shows many counties implement policy in this year

**Table 2.1:** CGSS and CHFS observations in estimations

| Timeline | Urban hukou: Meal Only | | Rural hukou: Meal + Subsidy | | All observations |
|---|---|---|---|---|---|
| CGSS | | | | | |
| | Untreated | Treated | Untreated | Treated | |
| Wave 2010 | 728 | 0 | 924 | 0 | 1652 |
| Wave 2011 | 335 | 0 | 579 | 0 | 914 |
| Wave 2012 | 512 | 175 | 582 | 256 | 1525 |
| Wave 2013 | 495 | 181 | 665 | 332 | 1673 |
| Wave 2015 | 280 | 133 | 404 | 217 | 1034 |
| Wave 2021 | 281 | 155 | 306 | 264 | 1006 |
| Sum | 2631 | 644 | 3460 | 1069 | 7804 |
| CHFS | | | | | |
| Timeline | Urban hukou: Meal Only | | Rural hukou: Meal + Subsidy | | All observations |
| | Untreated | Treated | Untreated | Treated | |
| Wave 1 (2011) | 1,414 | 0 | 2,153 | 0 | 3,567 |
| Wave 2 (2013) | 3,538 | 172 | 3,039 | 265 | 7,014 |
| Wave 3 (2015) | 2,736 | 217 | 3,377 | 312 | 6,642 |
| Wave 4 (2017) | 1,558 | 128 | 2,361 | 339 | 4,386 |
| Wave 5 (2019) | 1,829 | 41 | 2,176 | 129 | 4,175 |
| Sum | 11,075 | 558 | 13,106 | 1,045 | 25,784 |

**Note:** Data from CGSS 2010-2021 and CHFS 2011-2019. Treatment is defined in section 2.4

### 2.6.3 Variables

#### Outcome variables

We estimate the effects of the school meal policy on outcomes at the individual levels which includes the frequency of participation in various leisure activities, employment status, and weekly working hours. In the CGSS, the analysis explores how school meals influence parents' leisure time allocation. The survey includes questions on a wide range of activities such as watching TV or DVDs, going to the cinema, shopping, watching sports competitions in person, reading, participating in cultural activities, meeting with relatives and friends, listening to music, engaging in physical exercise, doing handicrafts, and using the internet. Respondents report how frequently they engage in each activity on a five-point scale from “every day” to “never,” and we convert these responses into continuous variables representing days per week spent on each activity.[23]

In our main analysis, treatment status may be correlated with the local availability of leisure activities. The construction of facilities such as cinemas, sports centers, and libraries is often closely tied to the level of local economic development and may share common funding sources with school canteen construction, potentially introducing endogeneity through local fiscal capacity. To mitigate this concern and ensure comparability across respondents, we focus on leisure activities that have near-universal availability regardless of regional economic differences. Specifically, we include watching television, shopping, socializing with relatives and friends, listening to music, physical exercise, and doing handicrafts.[24]

In the CHFS, we examine whether the school meal policy influences parents’ labour market behaviour. Employment status is captured as a binary variable equal to one if the respondent reports being employed. Weekly working hours are measured as a continuous variable, with non-employed individuals assigned a value of zero. Due to substantial non-response in the working hours variable, we apply non-response weights to ensure that individuals with valid data are representative of the broader employed population. These weights are constructed based solely on predetermined covariates, as information on occupation and industry is frequently missing.

We also explored the possibility of analysing intra-household labour allocation by calculating the difference between fathers’ and mothers’ weekly working hours. However, due to limited availability of matched parental data, the usable sample for this analysis is insufficient. Therefore, we instead present gender-specific labour outcomes separately.

### Controls

Our empirical specifications incorporate a rich set of control variables drawn from both the CGSS and CHFS datasets. At the individual level, we control for gender, age, educational attainment, whether co-residence with grandparents, whether working in a public sector, and

[23]The detail of coding leisure time activities is in the Appendix : Details of variables coding

[24]We list the availability of each activity around 2012, and we only include those activities with availability more than 90%. The detail of the availability of each activity is also in the Appendix : Details of leisure activities

the number of children in the household enrolled in preschool, compulsory education, and high school. To account for local socioeconomic context, we additionally control for county-level real per capita GDP in each year.[25]

### 2.6.4 Descriptive Statistics

Descriptive statistics from the CGSS and CHFS are presented in Table 2.2 and 2.3. We compare individuals across gender, educational attainment, and *hukou* status.

In terms of outcome variables, prior to the implementation of SNIP, television viewing frequency is broadly comparable between men and women. However, mothers are more likely to engage in leisure activities such as shopping, listening to music, and doing handicrafts. In contrast, fathers are more inclined to socialize with relatives and friends, exercise, and allocate more time to labour market participation. Educational disparities are also apparent: parents with less than a high school education tend to spend more time on passive leisure activities—such as television viewing and doing handicrafts—and less time in paid employment. In contrast, highly educated parents devote more time to meeting relatives and friends, shopping, listening to music, exercising, and participating in the labour market. Time allocation patterns of rural *hukou* parents are similar to those of less-educated urban *hukou* parents; they spend more time in watching TV and doing handicrafts than parents with urban *hukou*. However, there is no statistically significant difference in employment status between rural and urban *hukou* parents.

Regarding control variables, fathers are on average older than mothers and are less likely to be employed in the public sector. Highly educated parents are significantly more likely to have urban *hukou*, be employed in the public sector, and live in wealthier counties, compared to their less-educated counterparts. Similarly, rural *hukou* parents are less likely to work in the public sector and are more concentrated in economically disadvantaged areas.

[25] See Appendix : Details of Variable Coding for the construction of control variables.

**Table 2.2:** Descriptive statistics: CGSS

| | (1) Whole sample | (2) Mother | (3) Father | (4) Difference between Mother and Father | (5) High education | (6) Low education | (7) Difference between High and Low education | (8) Rural hukou | (9) Urban hukou | (10) Difference between rural and urban hukou |
|---|---|---|---|---|---|---|---|---|---|---|
| Children's availability of Meal only | 0.282 | 0.286 | 0.276 | 0.010 | 0.242 | 0.304 | -0.062*** | 0.000 | 0.218 | NA |
| | (0.450) | (0.452) | (0.447) | | (0.428) | (0.460) | | (0.000) | (0.413) | |
| Children's eligibility for Meal+Subsidy | 0.150 | 0.161 | 0.137 | 0.024*** | 0.052 | 0.206 | -0.154*** | 0.250 | 0.000 | NA |
| | (0.358) | (0.367) | (0.344) | | (0.222) | (0.404) | | (0.433) | (0.000) | |
| Whether have news about school canteen | 0.858 | 0.854 | 0.863 | -0.009 | 0.889 | 0.841 | 0.048*** | 0.854 | 0.863 | -0.009 |
| | (0.349) | (0.353) | (0.344) | | (0.315) | (0.366) | | (0.353) | (0.344) | |
| County's per capita GDP(in pounds) | 4270.000 | 4192.000 | 4368.000 | -176.000*** | 5247.000 | 3721.000 | 1526.000*** | 3784.000 | 5009.000 | -1225.000*** |
| | (3062.000) | (3044.000) | (3083.000) | | (3167.000) | (2859.000) | | (2860.000) | (3208.000) | |
| Outcomes: pre/untreated (times/week) | | | | | | | | | | |
| Watch TV | 4.591 | 4.593 | 4.590 | 0.003 | 4.316 | 4.760 | -0.444*** | 4.683 | 4.472 | 0.211*** |
| | (1.906) | (1.930) | (1.876) | | (2.004) | (1.823) | | (1.856) | (1.963) | |
| Go shopping | 1.670 | 1.939 | 1.333 | 0.606*** | 1.781 | 1.602 | 0.180*** | 1.574 | 1.797 | -0.223*** |
| | (1.319) | (1.330) | (1.226) | | (1.248) | (1.357) | | (1.334) | (1.290) | |
| Meet relatives and friends | 1.092 | 1.023 | 1.179 | -0.156*** | 1.301 | 0.964 | 0.337*** | 0.991 | 1.224 | -0.233*** |
| | (0.852) | (0.848) | (0.849) | | (0.822) | (0.845) | | (0.864) | (0.817) | |
| Listen music | 1.629 | 1.723 | 1.511 | 0.213*** | 2.136 | 1.319 | 0.817*** | 1.371 | 1.969 | -0.598*** |
| | (1.880) | (1.933) | (1.805) | | (1.875) | (1.815) | | (1.846) | (1.870) | |
| Exercise | 1.130 | 1.061 | 1.217 | -0.156*** | 1.763 | 0.744 | 1.020*** | 0.750 | 1.631 | -0.881*** |
| | (1.759) | (1.737) | (1.782) | | (1.923) | (1.527) | | (1.506) | (1.934) | |
| Doing handicrafts | 0.462 | 0.652 | 0.223 | 0.429*** | 0.388 | 0.507 | -0.119*** | 0.505 | 0.405 | 0.101*** |
| | (1.173) | (1.336) | (0.870) | | (1.035) | (1.247) | | (1.244) | (1.069) | |
| Controls | | | | | | | | | | |
| Urban hukou | 0.397 | 0.377 | 0.422 | -0.045*** | 0.744 | 0.202 | 0.543*** | 0.000 | 1.000 | NA |
| | (0.489) | (0.485) | (0.494) | | (0.436) | (0.401) | | (0.000) | (0.000) | |
| Age | 38.117 | 37.317 | 39.136 | -1.820*** | 37.939 | 38.217 | -0.279** | 37.856 | 38.514 | -0.659*** |
| | (5.451) | (5.393) | (5.355) | | (5.036) | (5.669) | | (5.641) | (5.124) | |
| Gender (male = 1) | 0.440 | 0.000 | 1.000 | NA | 0.496 | 0.408 | 0.088*** | 0.421 | 0.468 | -0.047*** |
| | (0.496) | (0.000) | (0.000) | | (0.500) | (0.492) | | (0.494) | (0.499) | |
| Whether living with grandparents | 0.209 | 0.188 | 0.235 | -0.048*** | 0.200 | 0.214 | -0.014 | 0.227 | 0.181 | 0.045*** |
| | (0.407) | (0.391) | (0.424) | | (0.400) | (0.410) | | (0.419) | (0.386) | |
| Whether working in the public sector | 0.150 | 0.128 | 0.178 | -0.050*** | 0.343 | 0.041 | 0.301*** | 0.042 | 0.313 | -0.271*** |
| | (0.357) | (0.334) | (0.382) | | (0.475) | (0.199) | | (0.201) | (0.464) | |
| Education level(in percentage) | | | | | | | | | | |
| no schooling | 0.047 | 0.068 | 0.020 | 0.048*** | 0.000 | 0.073 | NA | 0.070 | 0.010 | 0.060*** |
| | (-0.211) | (0.251) | (0.140) | | (0.000) | (0.260) | | (0.256) | (0.102) | |
| primary school | 0.206 | 0.240 | 0.164 | 0.076*** | 0.000 | 0.322 | NA | 0.306 | 0.056 | 0.250*** |
| | (-0.405) | (0.427) | (0.370) | | (0.000) | (0.467) | | (0.461) | (0.230) | |
| middle school | 0.388 | 0.369 | 0.411 | -0.041*** | 0.000 | 0.605 | NA | 0.472 | 0.259 | 0.212*** |
| | (0.487) | (0.483) | (0.492) | | (0.000) | (0.489) | | (0.499) | (0.438) | |
| high school | 0.182 | 0.163 | 0.207 | -0.044*** | 0.507 | 0.000 | NA | 0.118 | 0.279 | -0.161*** |
| | (0.386) | (0.369) | (0.405) | | (0.500) | (0.000) | | (0.323) | (0.449) | |
| college or higher | 0.177 | 0.161 | 0.198 | -0.037*** | 0.493 | 0.000 | NA | 0.034 | 0.395 | -0.361*** |
| | (0.382) | (0.367) | (0.399) | | (0.500) | (0.000) | | (0.181) | (0.489) | |
| Number of children in preschool | 0.144 | 0.152 | 0.133 | 0.020** | 0.090 | 0.174 | -0.084*** | 0.187 | 0.077 | 0.110*** |
| | (0.372) | (0.384) | (0.356) | | (0.297) | (0.406) | | (0.418) | (0.277) | |
| Number of children in primary school | 0.656 | 0.669 | 0.639 | 0.030** | 0.662 | 0.652 | 0.010 | 0.696 | 0.595 | 0.101*** |
| | (0.625) | (0.640) | (0.606) | | (0.604) | (0.637) | | (0.645) | (0.588) | |
| Number of children in middle school | 0.491 | 0.492 | 0.489 | 0.003 | 0.399 | 0.542 | -0.143*** | 0.507 | 0.466 | 0.041*** |
| | (0.592) | (0.601) | (0.582) | | (0.551) | (0.609) | | (0.610) | (0.563) | |
| N | 8,468 | 4,743 | 3,725 | | 3,044 | 5,424 | | 5,107 | 3,361 | |
| Number of counties | 108 | 108 | 108 | | 108 | 107 | | 106 | 108 | |

**Note:** Data from CGSS 2010-2021. Pooled sample. The total number of observations therefore reflects the size of the raw dataset, although the extent of missing values varies across variables. In headings, low education are lower than high school, and high education are high school or higher. Standard deviations in parentheses.

**Table 2.3:** Descriptive statistics: CHFS

| | (1) Whole sample | (2) Mother | (3) Father | (4) Difference between Mother and Father | (5) High education | (6) Low education | (7) Difference between High and Low education | (8) Rural hukou | (9) Urban hukou | (10) Difference between rural and urban hukou |
|---|---|---|---|---|---|---|---|---|---|---|
| Children's availability of Meal only | 0.343 | 0.343 | 0.344 | -0.002 | 0.277 | 0.374 | -0.097*** | 0.000 | 0.306 | NA |
| | (0.475) | (0.475) | (0.475) | | (0.448) | (0.484) | | (0.000) | (0.461) | |
| Children's eligibility for Meal+Subsidy | 0.182 | 0.184 | 0.181 | 0.002 | 0.055 | 0.242 | -0.186*** | 0.295 | 0.000 | NA |
| | (0.386) | (0.387) | (0.385) | | (0.229) | (0.428) | | (0.456) | (0.000) | |
| Whether have news about school canteen | 0.783 | 0.784 | 0.783 | 0.001 | 0.808 | 0.772 | 0.036*** | 0.758 | 0.825 | -0.068*** |
| | (0.412) | (0.411) | (0.412) | | (0.394) | (0.420) | | (0.429) | (0.380) | |
| County's per capita GDP(in pounds) | 4682.000 | 4697.000 | 4667.000 | 31.000 | 6300.000 | 3925.000 | 2375.000*** | 3927.000 | 5900.000 | -1973.000*** |
| | (4103.000) | (4116.000) | (4089.000) | | (5296.000) | (3129.000) | | (3071.000) | (5136.000) | |
| Outcomes: pre/untreated | | | | | | | | | | |
| Employment status | 0.843 | 0.754 | 0.933 | -0.179*** | 0.878 | 0.825 | 0.052*** | 0.844 | 0.842 | 0.002 |
| | (0.364) | (0.431) | (0.251) | | (0.328) | (0.380) | | (0.363) | (0.365) | |
| Weekly working hours | 38.051 | 30.678 | 46.048 | -15.369*** | 39.916 | 36.684 | 3.231*** | 38.247 | 37.833 | 0.414 |
| | (25.229) | (26.186) | (21.455) | | (20.608) | (28.060) | | (27.778) | (22.061) | |
| Controls | | | | | | | | | | |
| Urban hukou | 0.382 | 0.378 | 0.387 | -0.008* | 0.741 | 0.215 | 0.526*** | 0.000 | 1.000 | NA |
| | (0.486) | (0.485) | (0.487) | | (0.438) | (0.411) | | (0.000) | (0.000) | |
| Age | 38.740 | 37.826 | 39.655 | -1.829*** | 38.376 | 38.909 | -0.533*** | 38.725 | 38.763 | -0.037 |
| | (5.764) | (5.662) | (5.721) | | (5.108) | (6.039) | | (6.072) | (5.231) | |
| Gender (male = 1) | 0.500 | 0.000 | 1.000 | NA | 0.530 | 0.485 | 0.045*** | 0.496 | 0.505 | -0.009* |
| | (0.500) | (0.000) | (0.000) | | (0.499) | (0.500) | | (0.500) | (0.500) | |
| Whether living with grandparents | 0.232 | 0.229 | 0.235 | -0.006 | 0.193 | 0.251 | -0.058*** | 0.229 | 0.237 | -0.008** |
| | (0.422) | (0.420) | (0.424) | | (0.394) | (0.433) | | (0.420) | (0.425) | |
| Whether working in the public sector | 0.127 | 0.109 | 0.146 | -0.037*** | 0.307 | 0.043 | 0.264*** | 0.043 | 0.263 | -0.220*** |
| | (0.333) | (0.311) | (0.353) | | (0.461) | (0.203) | | (0.203) | (0.440) | |
| Education level(in percentage) | | | | | | | | | | |
| no schooling | 0.051 | 0.072 | 0.031 | 0.041*** | 0.000 | 0.075 | NA | 0.075 | 0.013 | 0.062*** |
| | (0.220) | (0.258) | (0.172) | | (0.000) | (0.263) | | (0.263) | (0.114) | |
| primary school | 0.228 | 0.250 | 0.206 | 0.044*** | 0.000 | 0.335 | NA | 0.316 | 0.086 | 0.229*** |
| | (0.420) | (0.433) | (0.405) | | (0.000) | (0.472) | | (0.465) | (0.281) | |
| middle school | 0.402 | 0.380 | 0.425 | -0.046*** | 0.000 | 0.591 | NA | 0.476 | 0.283 | 0.193*** |
| | (0.490) | (0.485) | (0.494) | | (0.000) | (0.492) | | (0.499) | (0.451) | |
| high school | 0.167 | 0.157 | 0.177 | -0.020*** | 0.524 | 0.000 | NA | 0.106 | 0.266 | -0.160*** |
| | (0.373) | (0.364) | (0.382) | | (0.500) | (0.000) | | (0.308) | (0.442) | |
| college or higher | 0.152 | 0.143 | 0.161 | -0.019*** | 0.476 | 0.000 | NA | 0.028 | 0.352 | -0.324*** |
| | (0.359) | (0.350) | (0.368) | | (0.500) | (0.000) | | (0.165) | (0.478) | |
| Number of children in preschool | 0.216 | 0.215 | 0.217 | -0.002 | 0.181 | 0.232 | -0.051*** | 0.252 | 0.158 | 0.094*** |
| | (0.455) | (0.454) | (0.456) | | (0.401) | (0.477) | | (0.489) | (0.387) | |
| Number of children in primary school | 0.734 | 0.731 | 0.736 | -0.005 | 0.709 | 0.745 | -0.036*** | 0.768 | 0.678 | 0.090*** |
| | (0.619) | (0.619) | (0.619) | | (0.540) | (0.652) | | (0.650) | (0.561) | |
| Number of children in middle school | 0.500 | 0.502 | 0.497 | 0.004 | 0.389 | 0.551 | -0.162*** | 0.534 | 0.444 | 0.090*** |
| | (0.569) | (0.570) | (0.569) | | (0.510) | (0.588) | | (0.590) | (0.530) | |
| N | 49,623 | 24,835 | 24,788 | | 15,816 | 33,807 | | 30,645 | 18,978 | |
| Number of individuals | 30,138 | 15,078 | 15,060 | | 10,628 | 19,510 | | 18,832 | 13,089 | |

**Note:** Source, CHFS: 2011-2019. Pooled sample. The total number of observations therefore reflects the size of the raw dataset, although the extent of missing values varies across variables. In headings, low education are lower than high school, and high education are high school or higher. Standard deviations in parentheses.

## 2.7 Methods

### 2.7.1 Regression and estimator

Our empirical analysis draws on two complementary data sources: county-level repeated cross-sectional data from the Chinese General Social Survey (CGSS) and individual-level panel data from the China Household Finance Survey (CHFS). Based on these two datasets, we estimate treatment effects separately for individuals with urban and rural *hukou* status.

We begin with a baseline specification implemented via a standard two-way fixed effects (TWFE) estimator:

$$Outcome_{ict} = \beta_1 Meal_{ct} + \beta_2 Subsidy_{ct} + \beta_3 X_{ict} + \gamma_c + \lambda_t + \epsilon_{ict} \tag{2.1}$$

where $Outcome_{ict}$ denotes the outcome variable for individual $i$ in county $c$ and year $t$. The outcome variables include leisure time allocation, employment status, and working hours. $Meal_{ct}$ is a binary indicator for counties implementing the school meal policy without subsidies (”Meal Only”), while $Subsidy_{ct}$ captures the presence of both school meals and targeted meal subsidies (”Meal + Subsidy”). By construction, $Meal_{ct} = 1$ whenever $Subsidy_{ct} = 1$. $\gamma_c$ and $\lambda_t$ denote county and year (or survey wave) fixed effects, respectively. $X_{ict}$ is a vector of individual-level covariates, and $\epsilon_{ict}$ is the idiosyncratic error term.

Although TWFE offers a convenient regression-based implementation of the Difference-in-Differences (DID) design, it implicitly uses later-treated units as controls for earlier-treated units. When treatment adoption is staggered, this feature can introduce bias if treatment effects are heterogeneous over time and across cohorts. Moreover, the TWFE residual variance and, consequently, the conventional clustered standard errors may be ”contaminated” by post-treatment treated units, leading to standard errors that are sensitive to sample composition in later periods and leads to bias. By contrast, the imputation method avoids these problems but relies on the untreated-period prediction model being sufficiently accurate, and misspecification of this model can itself lead to biased estimates.

To address these concerns, our preferred specification employs the imputation-based DID estimator of Borusyak et al. (2024), which separately identifies counterfactual outcomes for treated units and computes treatment effects without mixing treated and untreated observations across cohorts[26]. This method delivers unbiased estimates under parallel trends and produces period-specific standard errors that are not influenced by sample size fluctuations or outcome heterogeneity in other post-treatment periods.

Formally, the procedure consists of two steps (illustrated in Figure 2.2):

- **Imputation Step:** Estimate equation (2.1) on the donor sample $\Omega_0$, which includes all never-treated observations and treated observations in pre-treatment periods, excluding the treatment indicators. This step recovers county-*hukou* and time fixed effects as well as the coefficients on covariates.

- **Prediction and Estimation Step:** Using the estimated coefficients, time fixed effects and county fixed effects from the first step, we then predict untreated counterfactual outcomes for treated units in the post-treatment periods ($\Omega_1$ and $\Omega_2$). The treatment effect is computed as the mean difference between the observed outcome and its predicted counterfactual, interpreted as the intention-to-treat effect.

For individual-level analyses using CHFS panel data, we replace county fixed effects in the imputation step with individual fixed effects. This specification controls for time-invariant unobservables at the individual level and mitigates concerns about compositional changes at the county level. The trade-off is a reduction in effective sample size and the limited number of repeated observations per household.

[26]Alternative estimators for staggered treatment adoption include Callaway and Sant'Anna (2021) and Sun and Abraham (2021). We adopt the imputation method of Borusyak et al. (2024) for two reasons. First, both Callaway and Sant'Anna (2021) and Sun and Abraham (2021) are designed for panel data where the same individuals are tracked over time. Our primary data source, the CGSS, is a repeated cross-sectional survey clustered in the same counties each time, making cohort-level identification as required by these estimators infeasible. Second, Callaway and Sant'Anna (2021) estimate separate treatment effects for each cohort-period cell before aggregating, while Sun and Abraham (2021) rely on a large set of cohort-by-relative-time interaction terms. Both approaches become imprecise when the number of treated observations is limited, as is the case in our sample. The imputation estimator avoids these issues by pooling all untreated observations to recover treated unit (county) and time fixed effects, and then constructing counterfactual outcomes for each treated observation individually. This approach is efficient under limited treated sample sizes and extends naturally to our setting with two distinct treatments.

### 2.7.2 Identification Assumptions

Since this study employs a Difference-in-Differences (DID) imputation strategy to estimate the causal effects of the school meal policy, we need to outline the key identification assumptions underlying the DID framework.

As the analysis exploits variation in policy exposure across counties and demographic groups, several critical assumptions must hold for the DID estimates to be valid. Specifically, the following conditions are required:

- The treatment and control groups must exhibit parallel counterfactual trends in the absence of the policy.
- At the county (or county-*hukou*) level, individuals must not self-select into treatment based on anticipated policy exposure. That is, migration across counties should not be systematically driven by the school meal policy.
- The composition of the treatment and control groups must remain stable over time.

We formally test the parallel trends assumption prior to presenting the main results. In the sections below, we address other potential threats to identification, including unobserved heterogeneity across counties and the measurement errors.

### 2.7.3 Parallel Trends

To validate the key identification assumption of the Difference-in-Differences framework, we test whether treatment and control groups exhibited parallel trends in outcomes prior to the policy implementation. In the imputation approach, the control group $\Omega_0$ consists of units that were never treated as well as pre-treatment observations from treated units. All observations in $\Omega_0$ are used to test the parallel trends assumption.

Following Borusyak et al. (2024), we estimate the following event-study specification to assess pre-treatment dynamics:

$$Outcome_{ihct} = \sum_{t=-p}^{-1} \beta_t^{pre} \cdot COUNTDOWN_{ct} + \delta_2 X_{ihct} + \gamma_c + \lambda_t + \epsilon_{ihct} \tag{2.2}$$

In this specification, $COUNTDOWN_{ct}$ is a set of indicator variables that equal one if county $c$ is $t$ periods away from the policy implementation year (i.e., $t$ periods prior to treatment), and zero otherwise. Each $\beta_t^{pre}$ captures the average outcome differential between treated and control units in the $t^{th}$ pre-treatment period, relative to the omitted category. We also replace county fixed effect to individual fixed effect for testing employment status and weekly working hours.

The null hypotheses of parallel trends is tested by evaluating the joint significance of the pre-treatment coefficients using an F-test. Rejection of the null would suggest a violation of the parallel trends assumption. In addition to the joint test, we inspect individual $\beta_t^{pre}$ estimates and their confidence intervals to assess whether any anticipatory effects or differential trends exist prior to treatment.

This approach ensures that the treatment effect identified in subsequent regressions is not confounded by pre-existing trends in the outcome variable across treatment statuses.

### 2.7.4 Potential issues in identification

#### Macro-Level Heterogeneity

Systematic differences in baseline macroeconomic conditions across counties could confound our estimates if such characteristics are correlated with outcome trends. To address this concern, we examine macro-level indicators from the Sixth (2010) and Seventh (2020) National Population Censuses for early-treated, late-treated, and untreated counties. Details can be found in the Appendix Table A4 and A5.

The data from censuses indicate that treated counties were, on average, less economically developed relative to their untreated counterparts. To mitigate potential bias arising from differential economic trajectories, we include county-level per capita GDP as a time-varying control variable. Conditional on this control, we assume that residual macroeconomic heterogeneity does not bias our estimated treatment effects.

### Migration Dynamics

The policy targets students with rural *hukou*, and potential migration patterns that could threaten identification include: (1) from untreated/treated urban to treated rural areas; (2) from untreated rural to treated rural areas; and (3) from untreated urban to treated urban areas. However, these migrations are highly restricted under China's land tenure and *hukou* systems. Rural land is collectively owned and cannot be freely transferred to migrants, which means it is not possible to move from urban to rural area or get a rural *hukou*. Moreover, obtaining urban *hukou* in another county typically requires homeownership or prolonged employment, imposing significant costs. Given these institutional frictions, it is unlikely that school meal provision alone serves as a sufficient incentive for household relocation. Hence, migration is unlikely to be endogenous to the policy.

### Measurement Error

The first source of measurement error in our analysis arises from potential discrepancies between the *hukou* status of children and that of their parents. While the CGSS does not provide information on children's *hukou* status, data from the CHFS indicate that 89.32 percent of mothers and 93.49 percent of fathers share the same *hukou* type as their children. Since we categorize households into urban and rural groups based on parents' *hukou* status, this inconsistency may lead to misclassification of treatment status after the implementation of the policy.

Specifically, this type of error would not affect parents prior to treatment but could result in misidentifying "Meal Only" as "Meal + Subsidy," or vice versa, in the post-treatment period. For example, a small proportion of parents with urban *hukou* may be incorrectly classified as receiving "Meal Only" due to their child's rural *hukou*, and similarly, some parents with rural *hukou* may be incorrectly assigned to the subsidized group. This misclassification could lead to approximately 6-10 percent of the subsidy effect accounted into the urban sample, while simultaneously reducing the effect observed for rural *hukou* parents by a similar proportion. Given the comparable structure of the CGSS and CHFS samples, we expect the magnitude of this measurement error to be similar in both datasets.

Another source of measurement error arises from how we code counties to provide subsidized meals to poor students, as discussed in Section 2.4.2. This may introduce minor bias into the estimation. Based on national data, only 4.83% of working-age urban residents and 7.73% of rural residents received subsistence allowances in 2010—the official poverty line. Consequently, 256 (213) out of 644 (558) CGSS (CHFS) observations that treated as ”Meal Only” potentially also have subsidized meals on this bias, meaning that 1.9% (1.84%) of urban *hukou* respondents we classify as ”Meal Only” are likely actually to be eligible for a subsidized meal. Similarly, for students with rural *hukou*, 155 (373) out of 1,069 (1,045) CGSS (CHFS) observations that treated as ”Meal Only” potentially have subsidized meals, and 1.06% (2.76%) of rural respondents we classify as ”Meal Only” are likely actually to be eligible for a subsidized meal. These relatively small shares suggest that misclassification is unlikely to significantly bias our results. Furthermore, poverty status may vary over time, making fixed classification at the individual level infeasible.

## 2.8 Average impacts and heterogeneity by duration

In this section, we present the results of the parallel trends tests and estimate the average treatment effects for various outcome variables. Given that both the CGSS and CHFS datasets contain only five main survey years (or waves), we are limited to a maximum of three distinct pre-treatment periods. Observations from periods earlier than $t = -3$ are grouped into the $t = -3$ category to ensure balanced sample sizes across pre-treatment intervals.

Figure 2.3 and Figure 2.4 illustrate the dynamic treatment effects over time and capture the heterogeneity in treatment effects by duration of exposure. For the pre-treatment periods (prior to the policy implementation, indicated by the vertical red line), we employ the TWFE specification to test for parallel trends. Post-treatment estimates are derived using the imputation method discussed in the previous section.

Table 2.4 and Table 2.5 report the main estimation results. The first row in each table presents the average treatment effect of the school meal policy on the respective outcome. Below, we report the number of observations used in the regression and the number of treated units. The subsequent rows present the coefficients on the pre-treatment period indicators from the event-study specification (i.e., the ”countdown” coefficients). We also report the F-statistic and associated $p$-value from the joint test of the null hypotheses and find that most pre-treatment coefficients are jointly zero.

The F-tests and $p$-values in Table 2.4 and Table 2.5 indicate that the parallel trends assumption is violated in a limited number of cases. In particular, we reject the null of parallel pre-trends for urban leisure activities related to shopping, and this outcome is therefore excluded from the analysis.

### 2.8.1 Reallocation of time in leisure activities

Before presenting the average treatment effects, we first examine the impact of the policy on parents' participation in leisure activities before and after implementation, as illustrated in Figure 2.3.

We find an increase in the frequency of television viewing among rural *hukou* parents following the introduction of subsidized school meals, suggesting a potential reallocation of time from meal preparation to watching TV. Effect on the frequency of shopping precede the implementation of the policy for parents with urban *hukou*, but coefficients not significant different from zero for parents with rural *hukou*. Rural *hukou* parents exhibit a decline in shopping frequency two years after implementation. For the frequency of meeting with relatives and friends, urban *hukou* parents who received the "Meal Only" treatment show some positive and statistically significant changes. Meanwhile, rural *hukou* parents under the "Meal + Subsidy" condition exhibit negative and significant responses. However, the average level of the estimated coefficients remains largely unchanged from the pre-policy period. The key distinction lies in the standard errors, which are larger for the untreated and pre-treatment groups due to different estimation method. This pattern suggests that the observed statistical significance is primarily driven by estimation method rather than substantive behavioural changes.

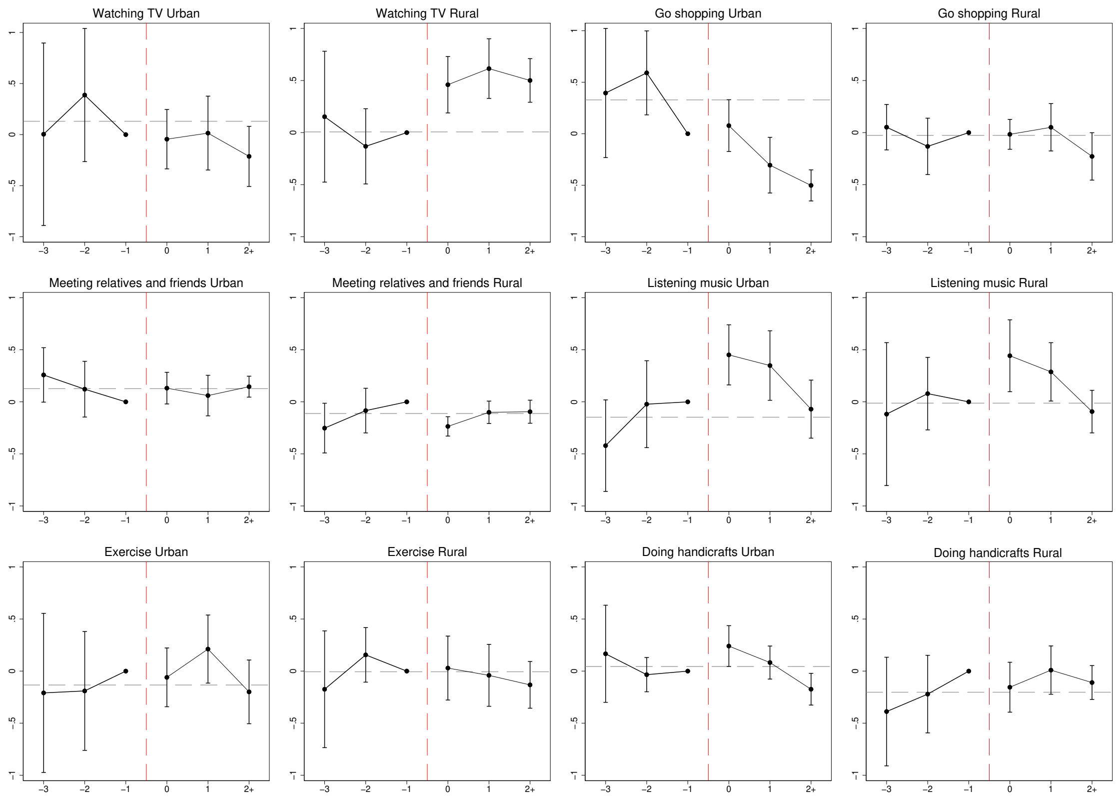
In the case of listening to music, both rural and urban *hukou* parents show an initial increase in participation following the policy implementation. This effect, however, becomes insignificant within two years, with frequencies returning to pre-policy levels. This indicates a lack of persistence in behavioural change. Finally, there is minimal variation in participation in exercise and doing handicrafts activities before and after the policy. This implies that time saved from reduced household cooking responsibilities was not reallocated to these two activities.

In Table 2.4, we list coefficients on the availability of school meals and subsidized meals for the frequency of different activities in leisure time. For each activity, the first column shows the effect of providing meal only limited in urban *hukou*, the second column lists the effect of meal subsidy limited in rural *hukou*.

In column (2), the provision of subsidized school meals to rural *hukou* parents is associated with a statistically significant increase in the frequency of television viewing, amounting to an additional 0.52 times per week. For shopping activities, the parallel trends assumption is not satisfied for parents with urban *hukou*, and the estimated average treatment effect for rural *hukou* parents is statistically insignificant. With respect to meeting with relatives and friends, both urban and rural *hukou* parents exhibit statistically significant changes following the policy, although in opposite directions. However, as mentioned above, these results are primarily driven by different method rather than a meaningful shift in the mean, suggesting that the policy effect may not reflect a substantive behavioural change. In columns (7) and (8), the "Meal Only" treatment increases the frequency of music listening among urban *hukou* parents by 0.19 times per week, while the "Meal + Subsidy" treatment increases the same activity among rural *hukou* parents by 0.15 times per week. For the remaining activities analysed, the school meal policy does not lead to any statistically significant changes in time allocation, implying that the policy did not influence parental behaviour in those activities.

In summary, the "Meal Only" implementation increases the frequency of music listening among parents with urban *hukou*, while the "Meal + Subsidy" implementation leads to higher frequencies of both television viewing and music listening among parents with rural *hukou*. These findings provide partial support for **hypotheses 1**. The school meal policy appears to reduce the time that rural *hukou* parents spend on home meal preparation, allowing them to reallocate time to different leisure activities, although not all types of leisure are affected. Furthermore, the observed increases in leisure activities such as television viewing and music listening may represent a potential welfare improvement for parents.

**Figure 2.3:** Effect of meal only and meal with subsidy on activities in leisure: before and after the implementation




**Note:** Data from CGSS 2010-2021. **Vertical axis is time of activities per week, and Horizontal axis is timeline counted in year.** Standard errors clustered by county and year. **TWFE method before implementation and Imputation method after implementation. The baseline is the mean of coefficients of COUNTDOWN variables:** $(\beta_{-1}^{pre} + \beta_{-2}^{pre} + \beta_{-3}^{pre})/3$ Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. Capped lines represent 90% confidence intervals. Because periods earlier than $t = -3$ are grouped into the $t = -3$ category, horizons are collinear and $-1$ is the base year.

**Table 2.4:** Treatment effects and parallel trends checks for impacts on school meal policy for frequency of activities in leisure time

| | Watch TV | | Go shopping | | Meeting relatives and friends | | Listening Music | | Exercise | | Doing handicrafts | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural |
| **Treatment effect** | -0.104 | 0.522*** | -0.281*** | -0.095 | 0.119** | -0.133*** | 0.192* | 0.152* | -0.050 | -0.065 | 0.014 | -0.088 |
| | (0.109) | (0.084) | (0.080) | (0.075) | (0.048) | (0.038) | (0.106) | (0.090) | (0.107) | (0.090) | (0.061) | (0.068) |
| Observations in pre-trends test regression | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 |
| Treated Observations | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 |
| Countdown Coefficients | | | | | | | | | | | | |
| Year horizon 1 | -0.004 | -0.153 | -0.395 | -0.052 | -0.258 | 0.252* | 0.421 | 0.118 | 0.209 | 0.174 | -0.166 | 0.389 |
| | (0.444) | (0.315) | (0.311) | (0.113) | (0.131) | (0.119) | (0.219) | (0.342) | (0.379) | (0.279) | (0.234) | (0.258) |
| Year horizon 2 | 0.384 | -0.285 | 0.196 | -0.184 | -0.137 | 0.168 | 0.398 | 0.197 | 0.019 | 0.330 | -0.200 | 0.167 |
| | (0.338) | (0.366) | (0.315) | (0.104) | (0.093) | (0.130) | (0.320) | (0.438) | (0.226) | (0.220) | (0.195) | (0.148) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations in pre-trends test regression | 2631 | 3457 | 2631 | 3457 | 2631 | 3457 | 2631 | 3457 | 2631 | 3457 | 2631 | 3457 |
| F-stat for all horizons | 1.22 | 0.37 | 4.27 | 1.57 | 2.36 | 2.23 | 1.90 | 0.12 | 0.23 | 2.28 | 0.75 | 1.14 |
| p(F) | 0.37 | 0.71 | 0.08 | 0.30 | 0.19 | 0.20 | 0.24 | 0.89 | 0.81 | 0.20 | 0.52 | 0.39 |

**Note:** Data from CGSS 2010-2021. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. "Meal_urban" is treated urban observation with meal only, and "Subsidy_rural" is treated rural observation with meal+subsidy. Year horizon 3 and earlier is the base year. Standard errors clustered by county and year in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

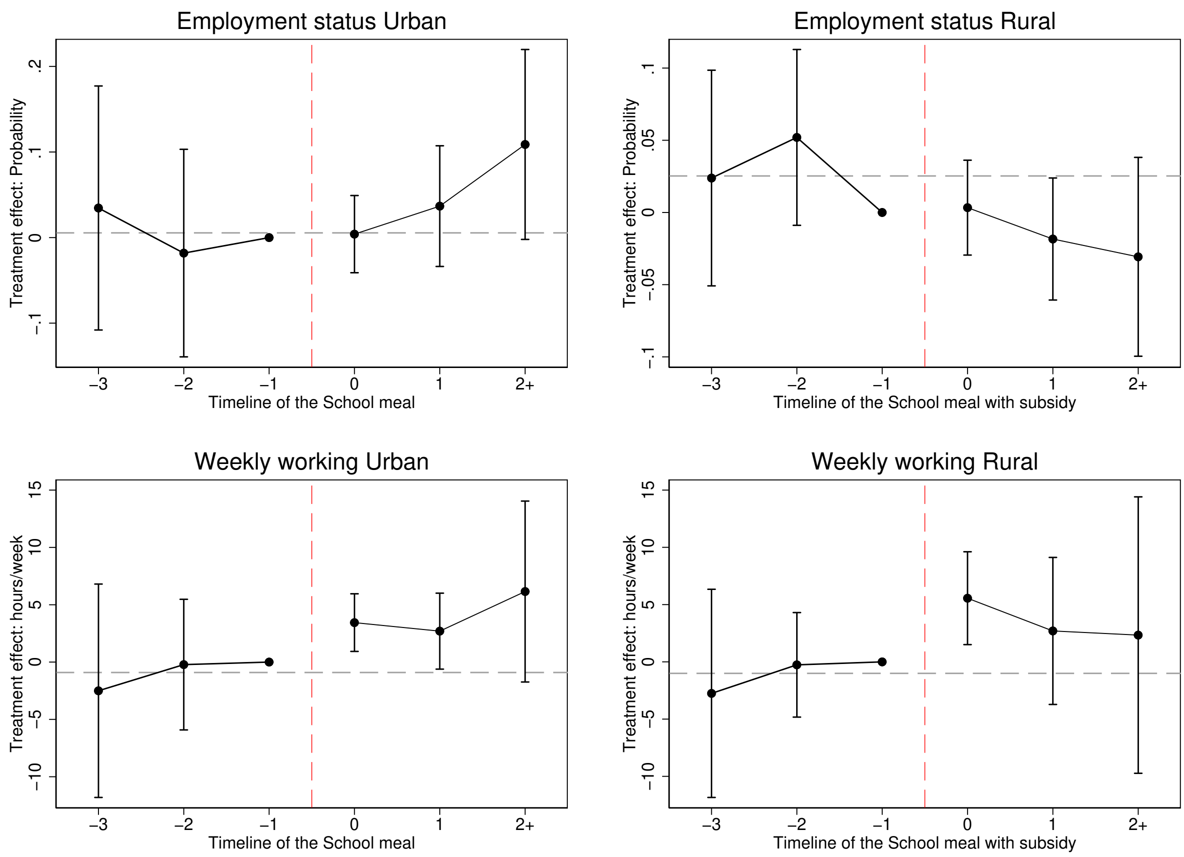
### 2.8.2 Reallocation of time in labour market

Figure 2.4 illustrates the dynamic effects of the school meal policy on parental labour market outcomes. Among urban *hukou* parents, employment status exhibits a significant increase following the implementation of the policy In contrast, rural *hukou* parents do not experience a statistically significant change in employment status after receiving subsidized meals. With respect to weekly working hours, both urban and rural *hukou* parents display an immediate increase upon the introduction of the subsidized meal program. This suggests that the policy may alleviate household time constraints, enabling parents to reallocate time from domestic responsibilities to paid labour.

Table 2.5 reports the average treatment effects of the school meal policy on parents' employment status and weekly working hours. The parallel trends assumption is not rejected, supporting the validity of the identification strategy. With respect to employment status, the estimated effects of both the ”Meal Only” and ”Meal + Subsidy” treatments are statistically insignificant, indicating no meaningful change in parents' likelihood of being employed. However, the policy does lead to a significant increase in weekly working hours. Specifically, urban *hukou* parents increase their working hours by approximately 3.49 hours per week, while rural *hukou* parents show an increase of about 3.79 hours.

The absence of significant effects on employment status, coupled with a substantial rise in working hours, suggests that the school meal policy primarily affects parents who are already employed. This implies that the policy may alleviate time constraints for working parents, thereby enabling more labour supply. In terms of magnitude, the estimates suggest that the policy increases parental working time by approximately 3.5 hours per week, or about 0.7 hours per working day. This magnitude lies between existing findings in the literature. For example, Holford and Rabe (2022) find that universal free school meals in England increase parents' working hours by around one hour per week, while G. Wang et al. (2024) show that China's Rural Nutrition Improvement Program reduces mothers' childcare time by about 2.8 hours per day. Given the relatively long time Chinese households typically spend on meal preparation, the magnitude of our estimates appears economically plausible.

**Figure 2.4:** Effect of meal only and meal with subsidy on labour outcomes: before and after implementation




**Note:** Data from CHFS 2011-2019. **Timeline count in Wave.** Standard errors clustered by county and year. **TWFE method before implementation and Imputation method after implementation. The baseline is the mean of COUNTDOWN variables:**$(\beta_{-1}^{pre} + \beta_{-2}^{pre} + \beta_{-3}^{pre})/3$ Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. Capped lines represent 90% confidence intervals. **1 = first survey wave since first exposed, where zero is the year of implementation.** Because periods earlier than $t = -3$ are grouped into the $t = -3$ category, horizons are collinear and $-1$ is the base year.

**Table 2.5:** Treatment effects and parallel trends checks for impacts on school meal policy for individual's labour outcomes

| | Employment status | | Weekly working hours | |
|---|---|---|---|---|
| | (1) Meal urban | (2) Subsidy rural | (3) Meal urban | (4) Subsidy rural |
| **Treatment effect** | 0.025 | -0.011 | 3.493*** | 3.787* |
| | (0.022) | (0.014) | (1.146) | (2.027) |
| Observations in pre-trends test regression | 5,667 | 7,751 | 4,322 | 3,153 |
| Treated Observations | 558 | 1,045 | 385 | 266 |
| **Countdown Coefficients** | | | | |
| Year horizon 1 | -0.035 | -0.031 | 2.495 | 2.316 |
| | (0.067) | (0.037) | (4.368) | (4.046) |
| Year horizon 2 | -0.053 | 0.023 | 2.277 | 2.207 |
| | (0.040) | (0.015) | (2.118) | (2.495) |
| Control | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Observations in pre-trends test regression | 5,667 | 7,751 | 4,322 | 3,153 |
| F-stat for all horizons | 0.88 | 3.91 | 0.92 | 0.58 |
| p(F) | 0.48 | 0.11 | 0.47 | 0.62 |

**Note:** Data from CHFS 2011-2019. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, family size, whether living with grandparents, whether work in public sector, school canteen dummy. "Meal urban" is treated urban observation with meal only, and "Subsidy rural" is treated rural observation with meal+subsidy. Year horizon 3 and earlier is the base year. Standard errors clustered by county and wave in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

## 2.9 Robustness checks for overall treatment effects

Table 2.6 and Table 2.7 present a series of robustness checks, including analyses based on samples that include divorced parents, regressions that not control for canteen information, and using a Two-Way Fixed Effects (TWFE) model.

First, we check how our sample selection affects the result. Divorced parents may face different time constraints compared to non-divorced parents. They often bear greater responsibilities in child rearing, which may lead to increased time spent in paid work. While the implementation of school meal programs can relieve some of their time burden, divorced parents may be more likely to reallocate the saved time toward labour market activities rather than leisure. However, our empirical results show that excluding or including divorced parents does not materially affect the estimates. Both the magnitude and statistical significance of the coefficients are similar. This suggests that, on average, the time allocation patterns of divorced parents are similar to those of other parents in the sample.

Second, in comparison with the existing literature, we incorporate additional information on school canteens, and this feature may also lead to differences in the results. Controlling for canteen information may absorb part of the effect of the school meal policy, as the estimated treatment effect then reflects the transition from the presence of a school canteen to the implementation of a universal meal program. Omitting this control could lead to an overestimation of the policy's impact by conflating the effects of existing infrastructure with those of the policy implementation. As shown in Table 2.6 and Table 2.7, some coefficients are indeed larger in the specifications that do not control for canteen information. However, the majority of the estimated coefficients and their statistical significance remain consistent across specifications, suggesting that the main results are robust to the inclusion of this control variable.

Third, results in our estimation could be different in different model specification, and we compare the Difference-in-Differences (DID) imputation method with the conventional two-way fixed effects (TWFE) model. Unlike TWFE, which estimates treatment effects by directly comparing treated and untreated groups over the same time periods, the DID imputation approach estimates the effect as the difference between observed outcomes and counterfactual outcomes predicted using pre-treatment trends. As a form of staggered DID, the imputation method is less sensitive to reductions in the number of control observations and is better suited to account for treatment effect heterogeneity across units and over time.

Compared with the baseline results obtained using the imputation method, the TWFE model produces some differences in coefficient magnitudes and statistical significance. In particular, the positive effect on weekly working hours for rural *hukou* parents becomes statistically insignificant, and the effects of the SNIP on employment status for rural *hukou* parents become significantly positive. These discrepancies are consistent with concerns in the literature that TWFE estimators may be biased when treatment effects are heterogeneous and treatment timing varies across units.

Last, given that we examine a total of 16 outcome variables, we address the issue of multiple hypotheses testing by applying the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Table 2.8 presents the adjusted p-values following the adjustment. Results are considered as significant and robust if the adjusted p-value is below 0.1, the p(F) value from the parallel trends test exceeds 0.1, and the estimated treatment effects differ from the pre-trend tests. The results show that weekly working hours for rural *hukou* parents and listening music for both urban and rural *hukou* parents parents are not significant. The robust estimates, which satisfy the parallel trends assumption, remain statistically significant after multiple hypothesis correction, and exhibit coefficients that differ from the pre-trend means, include increasing in frequency of watching TV among rural *hukou* parents as well as weekly working hours for parents with urban *hukou*.

**Table 2.6:** Robustness check: activities in leisure time

| | Watch TV | | Go shopping | | Meeting relatives and friends | | Listening Music | | Exercise | | Doing handicrafts | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural | Meal urban | Subsidy rural |
| **Baseline results** | -0.104 | 0.522*** | -0.281*** | -0.095 | 0.119** | -0.133*** | 0.192* | 0.152* | -0.050 | -0.065 | 0.014 | -0.088 |
| | (0.109) | (0.084) | (0.080) | (0.075) | (0.048) | (0.038) | (0.106) | (0.090) | (0.107) | (0.090) | (0.061) | (0.068) |
| Observations in underlying regression | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 |
| Treated Observations | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 |
| **With divorced parents** | -0.102 | 0.506*** | -0.278*** | -0.088 | 0.107** | -0.123*** | 0.167** | 0.172* | -0.058 | -0.040 | 0.035 | -0.067 |
| | (0.110) | (0.082) | (0.081) | (0.076) | (0.047) | (0.037) | (0.105) | (0.090) | (0.105) | (0.085) | (0.060) | (0.069) |
| Observations in underlying regression | 2,733 | 3,548 | 2,733 | 3,548 | 2,733 | 3,548 | 2,733 | 3,548 | 2,733 | 3,548 | 2,733 | 3,548 |
| Treated Observations | 671 | 1,099 | 671 | 1,099 | 671 | 1,099 | 671 | 1,099 | 671 | 1,099 | 671 | 1,099 |
| **Not control canteen information** | -0.101 | 0.525*** | -0.261*** | -0.081 | 0.133*** | -0.080** | 0.206* | 0.081 | 0.014 | -0.047 | -0.043 | -0.064 |
| | (0.109) | (0.084) | (0.079) | (0.076) | (0.047) | (0.037) | (0.112) | (0.087) | (0.105) | (0.090) | (0.061) | (0.068) |
| Observations in underlying regression | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 | 2,631 | 3,457 |
| Treated Observations | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 | 644 | 1,069 |
| **TWFE** | -0.061 | 0.496*** | -0.210 | -0.145 | 0.162** | -0.119 | 0.355** | 0.154 | -0.055 | 0.055 | 0.001 | -0.036 |
| | (0.197) | (0.124) | (0.138) | (0.094) | (0.071) | (0.073) | (0.155) | (0.160) | (0.169) | (0.184) | (0.078) | (0.119) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| N | 3,275 | 4,526 | 3,275 | 4,526 | 3,275 | 4,526 | 3,275 | 4,526 | 3,275 | 4,526 | 3,275 | 4,526 |

**Note:** Data from CGSS 2010-2021. Standard errors clustered by county and year in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$. "Meal_urban" is treated urban observation with meal only, and "Subsidy_rural" is treated rural observation with meal+subsidy.

**Table 2.7:** Robustness check: labour outcomes

| | Employment status | | Weekly working hours | |
|---|---|---|---|---|
| | (1) Meal urban | (2) Subsidy rural | (3) Meal urban | (4) Subsidy rural |
| **Original treatment effect** | 0.025 | -0.011 | 3.493*** | 3.787* |
| | (0.022) | (0.014) | (1.146) | (2.027) |
| Observations in underlying regression | 5,667 | 7,751 | 4,322 | 3,153 |
| Treated Observations | 558 | 1,045 | 385 | 266 |
| **With divorced parents** | 0.028 | -0.013 | 3.459*** | 3.700* |
| | (0.021) | (0.013) | (1.124) | (2.015) |
| Observations in underlying regression | 6,380 | 8,850 | 4,880 | 3,637 |
| Treated Observations | 568 | 1,061 | 393 | 268 |
| **Not control canteen information** | 0.029 | -0.006 | 3.590*** | 3.974* |
| | (0.022) | (0.014) | (1.194) | (2.010) |
| Observations in underlying regression | 5667 | 7751 | 4322 | 3153 |
| Treated Observations | 558 | 1,045 | 385 | 266 |
| **TWFE** | 0.102 | 0.048** | 2.410 | 0.283 |
| | (0.064) | (0.017) | (4.366) | (5.429) |
| Control | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| N | 6,225 | 8,796 | 4,707 | 3,419 |

**Note:** Data from CHFS 2011-2019. ”Meal_urban” is treated urban observation with meal only, and ”Subsidy_rural” is treated rural observation with meal+subsidy. Standard errors clustered by county and wave in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

**Table 2.8:** Robustness check: multiple hypothesis correction

| | Coefficient | unadjusted p-value | adjusted p-value | p(F) for parallel trends test | Treatment effects different with pretrend test | Significant and Robustness |
|---|---|---|---|---|---|---|
| **Urban (Meal Only)** | | | | | | |
| Watching TV | -0.104 | 0.343 | 0.457 | 0.370 | YES | NO |
| Go shopping | -0.281 | 0.001*** | 0.001*** | 0.083 | YES | NO |
| Meeting relatives and friends | 0.119 | 0.014** | 0.04** | 0.190 | NO | NO |
| Listening Music | 0.192 | 0.073* | 0.145 | 0.243 | YES | NO |
| Exercise | -0.050 | 0.642 | 0.733 | 0.806 | NO | NO |
| Doing DIY | 0.014 | 0.814 | 0.814 | 0.520 | NO | NO |
| Employment status | 0.025 | 0.253 | 0.405 | 0.482 | YES | NO |
| Weekly working hours | 3.493 | 0.003*** | 0.013** | 0.469 | YES | YES |
| **Rural (Meal + Subsidy)** | | | | | | |
| Watching TV | 0.522 | 0.000*** | 0.000*** | 0.710 | YES | YES |
| Go shopping | -0.095 | 0.209 | 0.279 | 0.295 | NO | NO |
| Meeting relatives and friends | -0.133 | 0.001*** | 0.003*** | 0.203 | NO | NO |
| Listening Music | 0.152 | 0.096* | 0.192 | 0.891 | YES | NO |
| Exercise | -0.065 | 0.471 | 0.4707 | 0.198 | NO | NO |
| Doing DIY | -0.088 | 0.203 | 0.325 | 0.392 | NO | NO |
| Employment status | -0.011 | 0.448 | 0.512 | 0.115 | YES | NO |
| Weekly working hours | 3.787 | 0.066* | 0.175 | 0.601 | YES | NO |

**Note:** Data from CGSS 2010-2021 and CHFS 2011-2019. Coefficients and unadjusted $p$-values are from the main outcomes reported in Section 2.8. Adjusted $p$-values account for multiple hypothesis testing using the Benjamini–Hochberg (BH) procedure. The BH method controls the false discovery rate (FDR) and adjusts each $p$-value $p(k)$ according to its rank $r_k$ among $M$ total hypotheses: $p_{\text{adj}}(k) = \min\left(1, \frac{p(k)\cdot M}{r_k}\right)$ where $r_k$ is the rank of the $k$th smallest $p$-value, and $M$ is the number of hypotheses tested (e.g., $M = 8$ for each group defined by *hukou* status). Adjusted $p$-values are capped at 1.

## 2.10 Heterogeneity by individual characteristics

In this section, we show the effects of school meals on parents with different education levels and gender. We divide observations into different groups by using their difference in characteristics[27], and we apply the DID imputation method in each subgroup. We only test heterogeneity of education on "Urban Meal", because very few of high educated parents have a rural *hukou*.

The parallel trends tests for heterogeneity are reported in Appendix A6–A10. We include in the analysis only those outcomes that satisfy the parallel trends assumption or those for which the estimated treatment effect moves in the opposite direction to the pre-treatment trend.[28]

[27] Low education is parents with middle school or lower education level; High education is parents with High school or higher education level. Gender includes male and female.

[28] An effect that moves in the opposite direction to the pre-treatment trend provides a lower bound for interpreting the treatment effect as non-zero. All outcomes are presented in the figures, and those excluded from the discussion are marked with an "X".

### 2.10.1 Policy effect on different education level

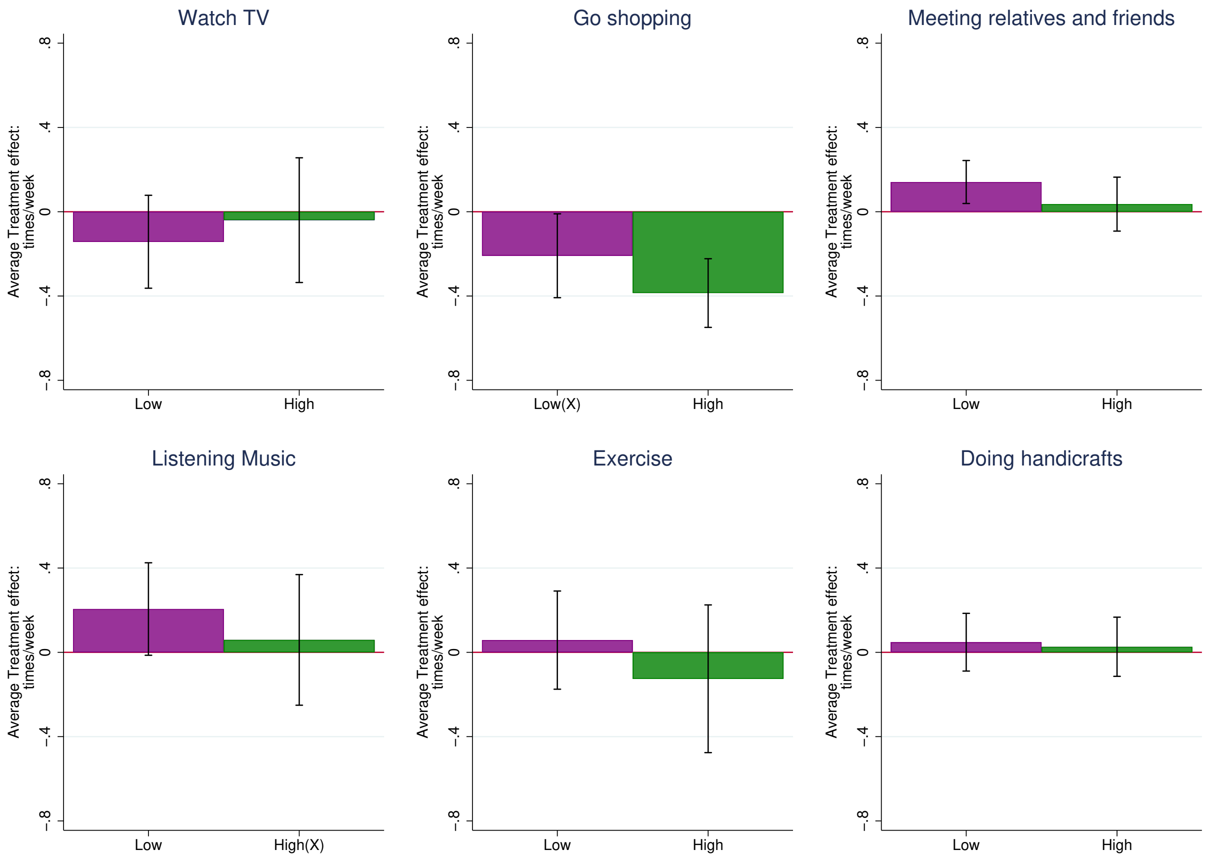
Parents with different earnings may have different restrictions in decisions, and their marginal utility of consumption and leisure could be different. However, the earnings are potentially endogenous to the treatment, and we need a predetermined measure for everyone. Education level is usually be considered as a powerful predictor of earnings. The heterogeneity of impact on parents with different education levels is shown in Figure 2.5 and Figure 2.6. The purple bar is for low educated parents in urban, and the green bar is for high educated parents in urban with capped lines representing 90% confidence intervals.

Figure 2.5 shows that, for both education groups, the policy does not lead to significant changes in time allocated to television viewing, listening music, exercise, or doing handicrafts activities. Turning to labour market outcomes in Figure 2.6, the increase in employment status among low-educated parents are not statistically significant, but the increase in weekly working hours is more significant.

As mentioned in framework, highly educated parents were assumed more likely to have their children eat in restaurants. Consistent with **hypotheses 3**, the meal (only) policy reduces very little of the cost of children's food, and this amount of saved money is so tiny compared with their earnings. The increase in school meal availability does not have any meaningful effect on their life, and we can not observe any significant effects on high-educated parents' employment status.

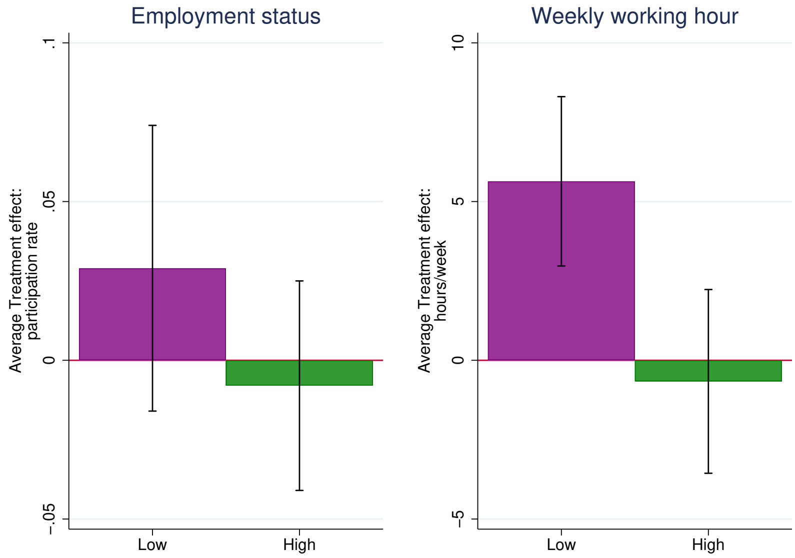
Parents with lower levels of education have been more likely to prepare lunch boxes for their children prior to the introduction of the school meal policy. When schools begin offering the "Meal Only" option, these parents face a choice between continuing home meal preparation and accepting the school meal. Referring back to **hypotheses 2**, access to school-provided meals potentially reduces the time associated with home cooking, thereby allowing parents to reallocate that time to other activities. This behavioural adjustment is partly consistent with our empirical findings, which show that low-educated parents experienced an increase in working hours following the implementation of the school meal policy.

**Figure 2.5:** Heterogeneity of impacts on school meal policy for activities in leisure time: By education level




**Note:** Data from CGSS 2010-2021. **Left bar is low educated parents in urban, and right bar is high educated parents in urban.** Standard errors clustered by county and year. Low education is lower than high school, and high education is high school or higher. Capped lines represent 90% confidence intervals. Regression details can be found in Appendix : Table A6. Results noted with "X" are those not satisfy the parallel trends assumption nor those for which the estimated treatment effect moves in the opposite direction to the pre-treatment trend.

**Figure 2.6:** Heterogeneity of impacts on school meal policy for labour outcomes: By education level




**Note:** Data from CHFS 2011-2019. **Left bar is low educated parents in urban, and right bar is high educated parents in urban.** Standard errors clustered by county and year. Low education is lower than high school, and high education is high school or higher. Capped lines represent 90% confidence intervals. Regression details can be found in Appendix : Table A7.

### 2.10.2 Policy effect on different gender

In a family, the mother and father may have different divisions of labour. If the mother spends more time making packed lunches, shopping, or doing other related work, the exposure to the school meal policy may have more impact on the mother than the father.

Figure 2.7 illustrates changes in mothers' and fathers' leisure activities following exposure to the "Meal Only" and "Meal + Subsidy" treatments. Both mothers and fathers exhibit similar responses in the frequency of television viewing, suggesting that this activity is likely undertaken jointly within the household. In contrast, gender differences emerge in many other leisure domains. Among rural *hukou* parents, mothers reduce their shopping frequency by 0.2 times per week after the policy implementation, whereas the effect on fathers is negligible. Regarding music listening, urban *hukou* fathers increase their participation following the implementation of school meals, while urban *hukou* mothers show no statistically significant change. Additional heterogeneity is observed in other leisure activities. Rural *hukou* mothers reduce their participation in exercise, whereas urban *hukou* fathers increase their engagement in doing handicrafts activities after the policy is introduced.

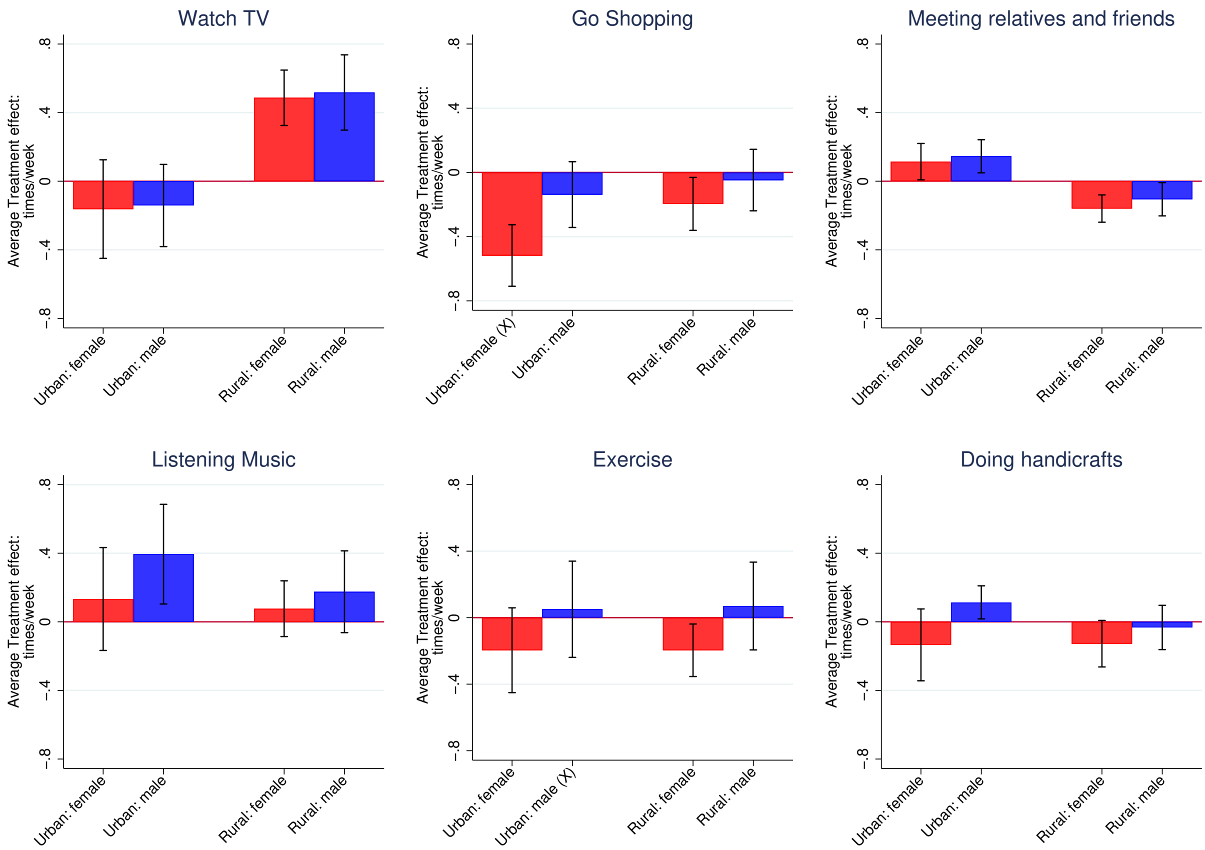
Figure 2.8 presents the heterogeneous effects of the school meal policy on parental labour outcomes by gender and *hukou* status. Among rural *hukou* mothers, employment status declines following the receipt of subsidized meals. In contrast, the other groups' increasing in employment status are insignificant. With respect to weekly working hours, urban *hukou* mothers experience a statistically significant increase after the implementation of the policy, suggesting that reducing time spent on housework allows for more labour supply. In contrast, rural *hukou* mothers show no significant change in working hours. Rural *hukou* fathers, however, exhibit an increase in time allocated to working following the introduction of subsidized meals.

Overall, the effects of the school meal policy differ by both gender and *hukou* status. An explanation for this heterogeneity lies in gender wage differentials across urban and rural households. Using data from the CHFS restricted to employed individuals with non-missing income information, we find that the average wage rate of urban *hukou* fathers is approximately 0.97 times that of urban *hukou* mothers, indicating near parity. In contrast, among rural *hukou* parents, the average wage rate for fathers is 1.37 times higher than that for mothers.

This disparity implies that the opportunity cost of time differs substantially between fathers and mothers in rural households. When the school meal policy reduces the need for time-intensive household tasks such as meal preparation, families may reallocate this freed time towards market work by the household member with the higher wage return. In rural households, where fathers earn considerably higher wages than mothers, the marginal return to additional labour supply is therefore greater for fathers. As a result, time savings generated by the policy are more likely to translate into increased working hours for rural fathers rather than for rural mothers. Consequently, we observe a greater increase in working hours for rural fathers compared to rural mothers following the implementation of the school meal policy.

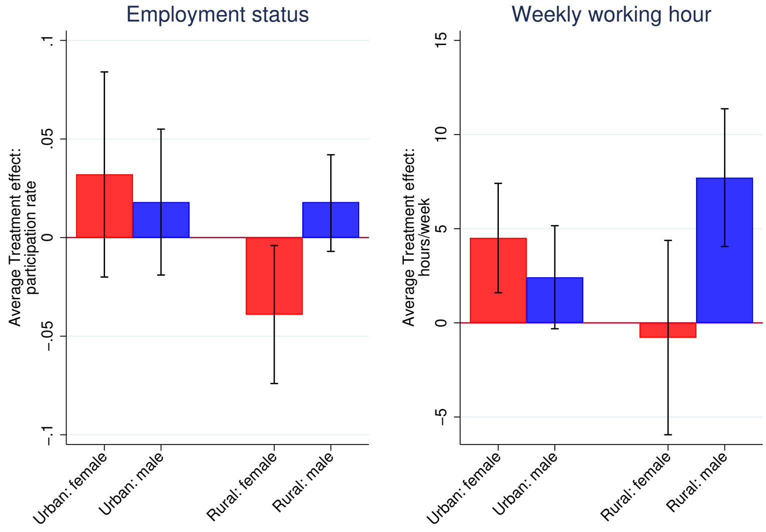
The decline in employment status among rural mothers further indicates that they are likely to undertake a larger share of housework. Since housework in rural households often requires greater physical effort than in urban settings, this may also explain why rural parents, on average, report less time spent on exercise prior to the SNIP (Table 2.2), as well as the observed reduction in exercise time among rural mothers after the program's introduction.

**Figure 2.7:** Heterogeneity of impacts on school meal policy for activities in leisure time: By gender




**Note:** Data from CGSS 2010-2021. **Left bar is female, and right bar is male.** Standard errors clustered by county and year. Low education is lower than high school, and high education is high school or higher. Capped lines represent 90% confidence intervals. Regression details can be found in Appendix : Table A8 and A9. Results noted with "X" are those not satisfy the parallel trends assumption nor those for which the estimated treatment effect moves in the opposite direction to the pre-treatment trend.

**Figure 2.8:** Heterogeneity of impacts on school meal policy for labour outcomes: By gender




**Note:** Data from CHFS 2011-2019. **Left bar is female, and right bar is male.** Standard errors clustered by county and year. Capped lines represent 90% confidence intervals. Regression details can be found in Appendix : Table A10.

## 2.11 Conclusion

This paper examines the indirect impact of school meal policy on parents. Specifically, we use the DID imputation method to test how implementing ”Meal Only” and ”Meal + Subsidy” affects parents’ leisure time allocation and labour market participation. We provide some evidence that the availability of meals (only) increases urban *hukou* parents’ time spent listening to music, and subsidized meals can increase rural *hukou* parents’ frequency of watching TV and listening to music. For labour market participation, we find that school meal policy significantly increase both urban and rural *hukou* parents’ time spent in working.

Consistent with the expectations outlined in our framework, urban *hukou* parents with lower levels of education appear to benefit more from time savings in home meal preparation. This group is more likely to increase both their working hours and participation in certain leisure activities following exposure to the school meal policy. In contrast, the effects on higher-educated parents are generally insignificant, suggesting that their time allocation is less constrained by household cooking. Regarding gender heterogeneity, as shown in Table 2.2

mothers exhibit a substantially higher baseline frequency of shopping, which approximately 40 percent greater than that of fathers. Therefore, it is unsurprising that school meal provision leads to a more pronounced reduction in shopping time for mothers. In addition, urban *hukou* mothers increase their working hours significantly after receiving the "Meal Only" treatment, consistent with the hypotheses that reduced domestic burdens can increase the labour supply. However, the effects of the "Meal + Subsidy" treatment exhibit notable asymmetries between genders within rural *hukou* households. Specifically, while the policy leads to a significant increase in weekly working hours for rural *hukou* fathers, it is associated with a decline in employment status among rural *hukou* mothers and has no significant effect on their weekly working hours. This difference is related to the gender wage gap within rural households. Since rural fathers earn substantially higher wages than rural mothers, the time saved through the school meal program is more efficiently reallocated to the labour market by fathers. This finding stands in contrast to much of the existing literature, which typically reports increases in mothers' labour supply following school meal implementations. Our results therefore suggest that the impact of school meal programs is shaped not only by gender but also by the gender wage gap.

Our results suggest that school meal programs targeting students can also have meaningful effects on their parents. These findings carry important implications for government policy. By providing school meals to children, the government can indirectly deliver welfare to working-age adults. Parents may experience reductions in time spent on household tasks such as meal preparation, allowing for increased leisure or labour market participation. In particular, school meal and meal subsidy may serve as an effective policy tool to boost labour supply, especially in low-earning households.

Our research has several limitations. First, the estimated effect reflects the expansion of school meal provision, yet we cannot figure out the proportion of pupils who were already receiving school meals before the SNIP. This is because detailed information on school canteen provision prior to the SNIP is unavailable, and we are therefore unable to assess the extent of measurement error. Second, the DID imputation method requires accurate prediction and therefore relies on retaining observations with sufficient pre-treatment periods. This restriction results in a relatively small treated sample, even though the overall datasets are large.

# References

Belot, M., & James, J. (2011). Healthy school meals and educational outcomes. *Journal of Health Economics*, *30*(3), 489–504.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bhattacharya, J., Currie, J., & Haider, S. J. (2006). Breakfast of champions? the school breakfast program and the nutrition of children and families. *Journal of Human Resources*, *41*(3), 445–466.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, *91*(6), 3253–3285.

Bütikofer, A., Mølland, E., & Salvanes, K. G. (2018). Childhood nutrition and labor market outcomes: Evidence from a school breakfast program. *Journal of Public Economics*, *168*, 62–80.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Chakraborty, T., & Jayaraman, R. (2019). School feeding and learning achievement: Evidence from india's midday meal program. *Journal of Development Economics*, *139*, 249–265.

Fang, G., & Zhu, Y. (2022). Long-term impacts of school nutrition: Evidence from china's school meal reform. *World Development*, *153*, 105854.

Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the school breakfast program. *Journal of Public Economics*, *124*, 91–104.

Holford, A., & Rabe, B. (2022). Going universal. the impact of free school lunches on child body weight outcomes. *Journal of Public Economics Plus*, *3*, 100016.

Holford, A., & Rabe, B. (2024). Universal free school meals and children's bodyweight. impacts by age and duration of exposure. *Journal of Health Economics*, *98*, 102937.

James, J. (2024). The effects of universal free school meals. *Wales Centre for Public Policy*.

Liang, Y., Chen, X., Zhao, C., & Jiang, S. (2022). Nutrition improvement program for rural compulsory education students and individual health. *Frontiers in Public Health*, *10*, 1051810.

Lundborg, P., Rooth, D.-O., & Alex-Petersen, J. (2022). Long-term effects of childhood nutrition: Evidence from a school lunch reform. *The Review of Economic Studies*, *89*(2), 876–908.

Marcus, M., & Yewell, K. G. (2022). The effect of free school meals on household food purchases: Evidence from the community eligibility provision. *Journal of Health Economics*, *84*, 102646.

Poppe, R., Frölich, M., & Haile, G. (2019). School meals and educational outcomes in rural ethiopia. *The Journal of Development Studies*, *55*(8), 1741–1756.

Rothbart, M. W., Schwartz, A. E., & Gutierrez, E. (2022). Paying for free lunch: The impact of cep universal free meals on revenues, spending, and student health. *Education Finance and Policy*, 1–46.

Schwartz, A. E., & Rothbart, M. W. (2020). Let them eat lunch: The impact of universal free meals on student performance. *Journal of Policy Analysis and Management*, *39*(2), 376–410.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Wang, G., Shi, X., & Golley, J. (2024). Feed the children, free the women? evidence from the china rural nutrition improvement program. *China Economic Review*, *87*, 102228.

Wang, H., & Cheng, Z. (2022). Kids eat free: School feeding and family spending on education. *Journal of Economic Behavior & Organization*, *193*, 196–212.

Wang, J., Zhou, L., & Yao, S. (2022). The impact of the nutrition improvement program on children's health in rural areas: Evidence from china. *Emerging Markets Finance and Trade*, *58*(1), 267–289.

Wei, C. e. a. (2019). Descriptive statistics of questionnaire data. In X. Zheng & C. Wei (Eds.), *Household energy consumption in china: 2016 report*. Springer.

Zheng, X., Ren, J., Chen, D., & Fang, X. (2022). School feeding and children's noncognitive skills: Evidence from the nutrition improvement program in rural china. *Applied Economics*, 1–20.

# Appendix

**Details of variables coding:**

In the CGSS questionnaire, the frequency of activities is reported on a 5-point scale, ranging from" Every day" to" Never" (Everyday = 1; Few times a week = 2; Few times a month = 3; Few times or less a year = 4; Never = 5). We convert each variable to a continuous days-per-week measure (Everyday = 6; Few times a week = 3; Few times a month = 2; Few times or less a year = 0.5; Never = 0).

Control variables in both CGSS and CHFS are coding as following : The evidence of the existence of school canteen ($0 = noevidence; 1 = existevidence$), gender ($0 = female; 1 = male$), age (in year), education level ($0 = noschooling; 1 = primaryschool; 2 = middleschool; 3 = highschoollevel; 4 = highereducation$), whether living with grandparents ($0 = no; 1 = yes$) and number of children in 3 different education levels (preschool, primary school, middle school).

**Details of leisure activities:**

Availability of watching TV is 97.6%; Resource from *National Bureau of Statistics*, available at: https://www.cnsaes.org.cn:8088/homepage/html/resource/res22/5822.html.

Availability of going to cinema is less than 21.2%; Resource from *2011 China Film Industry Research Report*, available at: https://cflac.org.cn/ysb/2011-06/08/content_22957000.htm.

Availability of going shopping is 100%, that people should always be available to local market.

Availability of reading books is 53.9%; Resource from *The Ninth National Reading Survey*, available at: https://culture.ifeng.com/whrd/detail_2012_04/20/14028668_0.shtml.

Availability of culture activities is around 20% to 40%; Resource from *Ministry of Culture and Tourism of the People's Republic of China*, available at: https://zwgk.mct.gov.cn/zfxxgkml/tjxx/202012/t20201204_906410.html.

Availability of meeting friends and relatives is 100%, that people should always be available to this activity.

Availability of listening to music is 96.8%; Resource from *National Bureau of Statistics*,

available at: https://www.cnsaes.org.cn:8088/homepage/html/resource/res22/5822.html.
Availability of exercise is 100%, that people should always be available to this activity.
Availability of opportunities to watch sports competitions is limited, particularly in less populated areas. Although we do not have precise statistics on coverage, competitive sports clubs such as those for football, basketball, and other major sports are typically located in cities with large populations. For example, the county with the smallest population that hosts a professional football club has approximately 210,000 residents. In contrast, smaller counties and rural areas rarely have access to such clubs or organized competitions, making sports viewing an activity that is not universally accessible across regions; Resource from *Chinese professional football league*, available at: https://www.csl-china.com/#/teamview
Availability of doing handicrafts is 100%, that people should always be available to this activity.
Availability of surfing Internet is 34%; Resource from *Xinhua News Agency*, available at: www.gov.cn/gzdt//2011-03/29/content_1833984.htm.

**Details of macro difference:**

Table A4 summarizes the Census of CGSS counties in 2010 and 2020. In the view of horizontal comparison, no matter whether in 2010 or 2020, counties that implemented school meal policies between 2010 and 2020 (later treated) have the lowest per capita GDP, the lowest urban population ratio, and the lowest average education year. These differences mean that the later treated counties have poorer economic environments than other counties. From the vertical view, the growth rate of per capita GDP in the later treated group is similar to that in the early treated groups, and later treated counties have the highest increase in urban population ratio and average education year. These data indicate that there is still a gap between treated counties and control counties, but they are closer in some aspects.

Table A5 presents the Census of CHFS counties in 2010 and 2020. Because many later treated counties whose treatment started in 2012 were not recorded in CHFS, the percentage of early treated counties is much higher in CHFS. This high percentage of early treated counties leads to similar macro statistics between early and later treated counties in 2010. Consistent with CGSS, never treated counties in CHFS also have a better economic environment than later treated counties, regardless of whether it is in 2010 or 2020. Overall, different groups of counties in CHFS have less macro difference than that in CGSS.

**Table A2.1:** Cumulative introduction of policy for CGSS counties (Number of counties: 108)

| Year | School meal from local government | | | | School meal from central government |
|---|---|---|---|---|---|
| | **Treatment: Meal only** Available to school meal[1] | Rural hukou boarding students eligible for subsidy | **Treatment: Meal + Subsidy** Rural hukou students eligible for subsidy[2] (local + central) | All students eligible for subsidy | Subsidy for rural hukou students |
| 2000 | 8 | **x** | **x** | **x** | **x** |
| 2001 | 8 | **x** | **x** | **x** | **x** |
| 2002 | 8 | **x** | **x** | **x** | **x** |
| 2003 | 8 | **x** | **x** | **x** | **x** |
| 2004 | 8 | **x** | **x** | **x** | **x** |
| 2005 | 13 | 5 | 5 | **x** | **x** |
| 2006 | 13 | 5 | 5 | **x** | **x** |
| 2007 | 13 | 5 | 5 | **x** | **x** |
| 2008 | 21 | 13 | 6 | **x** | **x** |
| 2009 | 25 | 17 | 6 | **x** | **x** |
| 2010 | 25 | 17 | 6 | **x** | **x** |
| 2011 | 26 | 18 | 6 | **x** | **x** |
| 2012 | 54 | 46 | 35 | 1 | 16 |
| 2013 | 54 | 46 | 35 | 1 | 16 |
| 2014 | 54 | 46 | 41 | 3 | 16 |
| 2015 | 57 | 49 | 44 | 3 | 16 |
| 2016 | 58 | 50 | 45 | 3 | 16 |
| 2017 | 58 | 50 | 45 | 3 | 16 |
| 2018 | 65 | 50 | 45 | 3 | 16 |
| 2019 | 68 | 53 | 48 | 3 | 16 |
| 2020 | 71 | 53 | 48 | 3 | 16 |
| 2021 | 73 | 54 | 49 | 4 | 16 |

[1] This column is the number of treated counties for ”availability of school meals.”
[2] This column is the number of treated counties for ”eligibility for subsidies.”

**Table A2.2:** Cumulative introduction of policy for CHFS counties (Number of counties: 257)

| Year | School meal from local government | | | | School meal from central government |
|---|---|---|---|---|---|
| | **Treatment: Meal only** Available to school meal[1] | Rural hukou boarding students eligible for subsidy | **Treatment: Meal + Subsidy** Rural hukou students eligible for subsidy[2] (local + central) | All students eligible for subsidy | Subsidy for rural hukou students |
| 2000 | 8 | **x** | **x** | **x** | **x** |
| 2001 | 8 | **x** | **x** | **x** | **x** |
| 2002 | 8 | **x** | **x** | **x** | **x** |
| 2003 | 8 | **x** | **x** | **x** | **x** |
| 2004 | 8 | **x** | **x** | **x** | **x** |
| 2005 | 19 | **x** | **x** | **x** | **x** |
| 2006 | 19 | **x** | **x** | **x** | **x** |
| 2007 | 29 | 10 | **x** | **x** | **x** |
| 2008 | 40 | 21 | 11 | **x** | **x** |
| 2009 | 40 | 21 | 11 | **x** | **x** |
| 2010 | 43 | 24 | 14 | **x** | **x** |
| 2011 | 43 | 24 | 14 | **x** | **x** |
| 2012 | 129 | 110 | 91 | **x** | 42 |
| 2013 | 134 | 115 | 96 | 1 | 42 |
| 2014 | 137 | 118 | 97 | 1 | 42 |
| 2015 | 139 | 120 | 99 | 2 | 42 |
| 2016 | 144 | 125 | 104 | 2 | 42 |
| 2017 | 149 | 130 | 109 | 2 | 42 |
| 2018 | 156 | 130 | 109 | 2 | 42 |
| 2019 | 168 | 142 | 120 | 2 | 42 |
| 2020 | 168 | 142 | 120 | 3 | 42 |
| 2021 | 172 | 145 | 123 | 4 | 42 |

[1] This column is the number of treated counties for ”availability of school meals.”
[2] This column is the number of treated counties for ”eligibility for subsidies.”

**Table A2.3:** Cumulative introduction of policy for CGSS and CHFS counties (Number of counties: 346)

| Year | School meal from local government | | | | School meal from central government |
|---|---|---|---|---|---|
| | **Treatment: Meal only** Available to school meal[1] | Rural hukou boarding students eligible for subsidy | **Treatment: Meal + Subsidy** Rural hukou students eligible for subsidy[2] (local + central) | All students eligible for subsidy | Subsidy for rural hukou students |
| 2000 | 12 | x | x | x | x |
| 2001 | 12 | x | x | x | x |
| 2002 | 12 | x | x | x | x |
| 2003 | 12 | x | x | x | x |
| 2004 | 12 | x | x | x | x |
| 2005 | 27 | 5 | 5 | x | x |
| 2006 | 27 | 5 | 5 | x | x |
| 2007 | 37 | 15 | 5 | x | x |
| 2008 | 55 | 33 | 17 | x | x |
| 2009 | 59 | 37 | 17 | x | x |
| 2010 | 62 | 40 | 20 | x | x |
| 2011 | 63 | 41 | 20 | x | x |
| 2012 | 173 | 151 | 121 | 1 | 57 |
| 2013 | 178 | 156 | 126 | 2 | 57 |
| 2014 | 181 | 159 | 133 | 3 | 57 |
| 2015 | 186 | 164 | 138 | 4 | 57 |
| 2016 | 192 | 170 | 144 | 4 | 57 |
| 2017 | 197 | 175 | 149 | 4 | 57 |
| 2018 | 207 | 175 | 149 | 4 | 57 |
| 2019 | 222 | 190 | 163 | 4 | 57 |
| 2020 | 225 | 190 | 163 | 5 | 57 |
| 2021 | 230 | 193 | 166 | 6 | 57 |

[1] This column is the number of treated counties for ”availability of school meals.”
[2] This column is the number of treated counties for ”eligibility for subsidies.”

**Table A2.4:** Summary of Census (CGSS counties)

| Census information in 2010 | Early treated (24 counties) | | Later treated (43 counties) | | Never treated (41 counties) | |
|---|---|---|---|---|---|---|
| | Mean | sd | Mean | sd | Mean | sd |
| Per capita GDP(in pounds) | 5593.13 | 2909.41 | 3075.14 | 2635.24 | 3946.64 | 3277.45 |
| Sex ratio(100 female) | 106.54 | 5.68 | 104.83 | 6.67 | 104.11 | 5.39 |
| Urban population ratio(%) | 67.11% | 0.28 | 46.05% | 0.26 | 63.06% | 0.27 |
| Married ratio in population over 15(%) | 70.95% | 0.05 | 72.37% | 0.03 | 69.70% | 0.07 |
| Average number of surviving child(per person) | 1.12 | 0.35 | 1.39 | 0.30 | 1.16 | 0.31 |
| Employed ratio in population over 16(%) | 64.44% | 0.09 | 70.39% | 0.09 | 62.56 | 0.12 |
| High school or higher education population ratio(%) | 38.38% | 0.18 | 23.38% | 0.15 | 33.87% | 0.18 |
| Male's average education year | 10.48 | 1.47 | 9.32 | 1.31 | 10.18 | 1.37 |
| Female's average education year | 9.82 | 1.83 | 8.37 | 1.53 | 9.58 | 1.58 |
| Total average education year | 10.16 | 1.64 | 8.85 | 1.41 | 9.88 | 1.47 |

| Census information in 2020 | Early treated (24 counties) | | Later treated (43 counties) | | Never treated (41 counties) | |
|---|---|---|---|---|---|---|
| | Mean | sd | Mean | sd | Mean | sd |
| Per capita GDP(in pounds) | 9719.17 | 4240.25 | 5340.80 | 3118.30 | 5563.28 | 2964.00 |
| Sex ratio(100 female) | 104.82 | 6.62 | 103.45 | 3.61 | 102.69 | 6.21 |
| Urban population ratio(%) | 75.09% | 0.22 | 57.16% | 0.21 | 72.34% | 0.23 |
| Married ratio in population over 15(%) | 74.14% | 0.03 | 73.97% | 0.04 | 71.01% | 0.06 |
| Average number of surviving child(per person) | 1.14 | 0.31 | 1.37 | 0.30 | 1.12 | 0.26 |
| Employed ratio in population over 16(%) | 58.43% | 0.06 | 59.96% | 0.08 | 54.11% | 0.10 |
| High school or higher education population ratio(%) | 44.42% | 0.19 | 30.12% | 0.15 | 40.06% | 0.17 |
| Male's average education year | 10.72 | 1.55 | 9.57 | 1.27 | 10.34 | 1.27 |
| Female's average education year | 10.29 | 1.87 | 8.99 | 1.46 | 10.00 | 1.48 |
| Total average education year | 10.51 | 1.70 | 9.28 | 1.36 | 10.17 | 1.37 |

**Note:** Data from census 2010 and 2020. Unweighted. Pooled sample with all CGSS counties in the regressions. Available at: https://www.stats.gov.cn/zt_18555/zdtjgz/zgrkpc/

**Table A2.5:** Summary of Census (CHFS counties)

| Census information in 2010 | Early treated (100 counties) | | Later treated (57 counties) | | Never treated (100 counties) | |
|---|---|---|---|---|---|---|
| | Mean | sd | Mean | sd | Mean | sd |
| Per capita GDP(in pounds) | 3443.58 | 2433.17 | 3225.26 | 3611.33 | 3460.63 | 2319.70 |
| Sex ratio(100 female) | 104.87 | 4.63 | 105.23 | 7.70 | 104.40 | 5.02 |
| Urban population ratio(%) | 53.40% | 0.28 | 51.56% | 0.30 | 61.36% | 0.28 |
| Married ratio in population over 15(%) | 70.88% | 0.06 | 72.42% | 0.04 | 69.54% | 0.06 |
| Average number of surviving child(per person) | 1.32 | 0.35 | 1.30 | 0.30 | 1.24 | 0.33 |
| Employed ratio in population over 16(%) | 68.07% | 0.08 | 67.31% | 0.11 | 63.85% | 0.10 |
| High school or higher education population ratio(%) | 26.76% | 0.17 | 25.96% | 0.16 | 32.37% | 0.18 |
| Male's average education year | 9.46 | 1.63 | 9.49 | 1.27 | 10.14 | 1.35 |
| Female's average education year | 8.58 | 1.91 | 8.76 | 1.54 | 9.43 | 1.54 |
| Total average education year | 9.02 | 1.76 | 9.13 | 1.40 | 9.79 | 1.44 |

| Census information in 2020 | Early treated (100 counties) | | Later treated (57 counties) | | Never treated (100 counties) | |
|---|---|---|---|---|---|---|
| | Mean | sd | Mean | sd | Mean | sd |
| Per capita GDP(in pounds) | 5985.12 | 3831.95 | 4617.07 | 3752.75 | 5266.73 | 3100.09 |
| Sex ratio(100 female) | 104.52 | 5.74 | 104.24 | 5.89 | 103.23 | 5.37 |
| Urban population ratio(%) | 63.58% | 0.23 | 60.62% | 0.24 | 69.13% | 0.24 |
| Married ratio in population over 15(%) | 72.49% | 0.05 | 74.30% | 0.03 | 71.56% | 0.05 |
| Average number of surviving child(per person) | 1.29 | 0.32 | 1.29 | 0.28 | 1.23 | 0.33 |
| Employed ratio in population over 16(%) | 57.51% | 0.08 | 58.01% | 0.11 | 54.46% | 0.08 |
| High school or higher education population ratio(%) | 33.71% | 0.17 | 31.23% | 0.15 | 38.09% | 0.17 |
| Male's average education year | 9.76 | 1.54 | 9.62 | 1.22 | 10.24 | 1.29 |
| Female's average education year | 9.17 | 1.79 | 9.16 | 1.42 | 9.82 | 1.48 |
| Total average education year | 9.47 | 1.66 | 9.40 | 1.32 | 10.03 | 1.37 |

**Note:** Data from census 2010 and 2020. Unweighted. Pooled sample with all CHFS counties in the regressions. Available at:https://www.stats.gov.cn/zt_18555/zdtjgz/zgrkpc/

**Table A2.6:** Heterogeneity of impacts on school meal policy for activities in leisure time: By education level

| | Watch TV | | Go shopping | | Meeting relatives and friends | | Listening Music | | Exercise | | Doing DIY | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Urban low | Urban high | Urban low | Urban high | Urban low | Urban high | Urban low | Urban high | Urban low | Urban high | Urban low | Urban high |
| **Treatment effect** | -0.143 | -0.040 | -0.209* | -0.386*** | 0.141** | 0.036 | 0.206 | 0.059 | 0.058 | -0.126 | 0.048 | 0.026 |
| | (0.133) | (0.178) | (0.120) | (0.098) | (0.061) | (0.077) | (0.132) | (0.186) | (0.140) | (0.210) | (0.082) | (0.084) |
| Observations in pre-trends test regression | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 |
| Treated Observations | 369 | 259 | 369 | 259 | 369 | 259 | 369 | 259 | 369 | 259 | 369 | 259 |
| Countdown Coefficients | | | | | | | | | | | | |
| Year horizon 1 | -0.071 | 0.022 | -0.653 | 0.068 | -0.243 | -0.277 | 0.226 | 0.766* | 0.045 | 0.843 | -0.164 | -0.182 |
| | (0.451) | (0.670) | (0.418) | (0.270) | (0.129) | (0.267) | (0.318) | (0.363) | (0.356) | (0.473) | (0.267) | (0.323) |
| Year horizon 2 | 0.440 | 0.189 | 0.356 | 0.004 | -0.055 | -0.224 | 0.538* | 0.059 | -0.026 | 0.510 | -0.171 | -0.243 |
| | (0.474) | (0.654) | (0.446) | (0.235) | (0.135) | (0.146) | (0.223) | (0.706) | (0.356) | (0.307) | (0.262) | (0.289) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations in pre-trends test regression | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 | 1,586 | 1,033 |
| F-stat for all horizons | 0.90 | 0.11 | 6.83 | 0.06 | 1.79 | 1.18 | 3.15 | 5.42 | 0.02 | 1.90 | 0.22 | 0.45 |
| p(F) | 0.46 | 0.90 | 0.04 | 0.94 | 0.26 | 0.38 | 0.13 | 0.06 | 0.99 | 0.24 | 0.81 | 0.66 |

**Note:** Data from CGSS 2010-2021. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. Year horizon 3 is the base year. Standard errors clustered by county and year in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

**Table A2.7:** Heterogeneity of impacts on school meal policy for labour outcomes: By education level

| | Employment status | | Weekly working hours | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Urban low | Urban high | Urban low | Urban high |
| **Treatment effect** | 0.029 | -0.008 | 5.638*** | -0.664 |
| | (0.027) | (0.019) | (1.599) | (1.721) |
| Observations in pre-trends test regression | 3,563 | 2,104 | 2,517 | 1,805 |
| Treated Observations | 414 | 144 | 254 | 131 |
| **Countdown Coefficients** | | | | |
| Year horizon 1 | -0.060 | 0.036 | 3.781 | 0.882 |
| | (0.078) | (0.084) | (5.422) | (4.630) |
| Year horizon 2 | -0.098 | 0.035 | 1.121 | 3.700 |
| | (0.060) | (0.041) | (3.230) | (1.900) |
| Control | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Observations in pre-trends test regression | 3,563 | 2,104 | 2,517 | 1,805 |
| F-stat for all horizons | 1.36 | 0.48 | 0.25 | 3.57 |
| p(F) | 0.35 | 0.65 | 0.79 | 0.13 |

**Note:** Data from CHFS 2011-2019. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, family size, whether living with grandparents, whether work in public sector, school canteen dummy. Year horizon 3 is the base year. Standard errors clustered by county and wave in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

**Table A2.8:** Heterogeneity of impacts on school meal policy for activities in leisure time: By gender

| | Watch TV | | | | Go shopping | | | | Meeting relatives and friends | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Urban female | Urban male | Rural female | Rural male | Urban female | Urban male | Rural female | Rural male | Urban female | Urban male | Rural female | Rural male |
| **Treatment effect** | -0.163 | -0.141 | 0.487*** | 0.518*** | -0.518*** | -0.138 | -0.196* | -0.048 | 0.114* | 0.146** | -0.160*** | -0.105* |
| | (0.173) | (0.144) | (0.097) | (0.132) | (0.115) | (0.123) | (0.099) | (0.115) | (0.064) | (0.058) | (0.048) | (0.058) |
| Observations in pre-trends test regression | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 |
| Treated Observations | 329 | 301 | 647 | 384 | 329 | 301 | 647 | 384 | 329 | 301 | 647 | 384 |
| Countdown Coefficients | | | | | | | | | | | | |
| Year horizon 1 | 0.213 | -0.022 | -0.248 | 0.148 | -0.557 | -0.038 | -0.074 | -0.010 | -0.438** | -0.068 | 0.166 | 0.265 |
| | (0.381) | (0.738) | (0.294) | (0.423) | (0.458) | (0.325) | (0.148) | (0.343) | (0.150) | (0.269) | (0.154) | (0.154) |
| Year horizon 2 | 0.781* | 0.035 | -0.522 | 0.002 | 0.495 | -0.017 | -0.253* | -0.107 | -0.256** | 0.002 | 0.224 | 0.069 |
| | (0.319) | (0.568) | (0.341) | (0.450) | (0.421) | (0.323) | (0.104) | (0.194) | (0.086) | (0.183) | (0.169) | (0.171) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations in pre-trends test regression | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 |
| F-stat for all horizons | 4.73 | 0.01 | 1.29 | 0.24 | 5.17 | 0.01 | 3.09 | 0.29 | 6.61 | 0.04 | 0.91 | 2.63 |
| p(F) | 0.07 | 0.99 | 0.35 | 0.79 | 0.06 | 0.99 | 0.13 | 0.76 | 0.04 | 0.96 | 0.46 | 0.17 |

**Note:** Data from CGSS 2010-2021. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. Year horizon 3 is the base year. Standard errors clustered by county and year in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

**Table A2.9:** Heterogeneity of impacts on school meal policy for activities in leisure time: By gender : continued

| | Listening Music | | | | Exercise | | | | Doing DIY | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) | (22) | (23) | (24) |
| | Urban female | Urban male | Rural female | Rural male | Urban female | Urban male | Rural female | Rural male | Urban female | Urban male | Rural female | Rural male |
| **Treatment effect** | 0.133 | 0.395** | 0.076 | 0.175 | -0.196 | 0.051 | -0.196** | 0.070 | -0.135 | 0.113* | -0.128 | -0.033 |
| | (0.181) | (0.175) | (0.097) | (0.143) | (0.154) | (0.175) | (0.095) | (0.158) | (0.126) | (0.058) | (0.081) | (0.077) |
| Observations in pre-trends test regression | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 |
| Treated Observations | 329 | 301 | 647 | 384 | 329 | 301 | 647 | 384 | 329 | 301 | 647 | 384 |
| Countdown Coefficients | | | | | | | | | | | | |
| Year horizon 1 | 0.849 | 0.487 | 0.292 | -0.103 | 0.166 | 0.301 | 0.153 | 0.172 | -0.677 | 0.055 | 0.340 | 0.466 |
| | (0.510) | (0.438) | (0.517) | (0.329) | (0.587) | (0.554) | (0.280) | (0.372) | (0.404) | (0.142) | (0.349) | (0.261) |
| Year horizon 2 | 0.836 | 0.161 | 0.331 | -0.080 | 0.370 | -0.437 | 0.325 | 0.317 | -0.560 | -0.035 | 0.252 | -0.007 |
| | (0.448) | (0.438) | (0.590) | (0.426) | (0.572) | (0.241) | (0.268) | (0.303) | (0.363) | (0.091) | (0.271) | (0.126) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations in pre-trends test regression | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 | 1,413 | 1,208 | 1,978 | 1,475 |
| F-stat for all horizons | 1.78 | 0.72 | 0.17 | 0.08 | 0.32 | 6.82 | 0.98 | 0.61 | 1.41 | 0.67 | 0.53 | 4.06 |
| p(F) | 0.26 | 0.53 | 0.85 | 0.92 | 0.74 | 0.04 | 0.44 | 0.58 | 0.33 | 0.55 | 0.62 | 0.09 |

**Note:** Data from CGSS 2010-2021. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, whether living with grandparents, whether work in public sector, school canteen dummy. Year horizon 3 is the base year. Standard errors clustered by county and year in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

**Table A2.10:** Heterogeneity of impacts on school meal policy for labour outcomes: By gender

| | Employment status | | | | Weekly working hours | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Urban female | Urban male | Rural female | Rural male | Urban female | Urban male | Rural female | Rural male |
| **Treatment effect** | 0.032 | 0.018 | -0.039* | 0.018 | 4.503** | 2.422 | -0.782 | 7.712*** |
| | (0.031) | (0.022) | (0.021) | (0.015) | (1.743 ) | (1.640) | (3.085) | (2.188) |
| Observations in pre-trends test regression | 2,802 | 2,865 | 3,923 | 3,828 | 2,236 | 2,086 | 1,667 | 1,486 |
| Treated Observations | 414 | 144 | 414 | 144 | 254 | 131 | 254 | 131 |
| **Countdown Coefficients** | | | | | | | | |
| Year horizon 1 | -0.049 | -0.027 | -0.050 | -0.013 | -1.594 | 5.982 | 1.756 | 2.346 |
| | (0.109) | (0.055) | (0.047) | (0.041) | (4.892) | (4.533) | (5.734) | (3.896) |
| Year horizon 2 | -0.035 | -0.071 | 0.021 | 0.024 | 2.700 | 1.606 | 1.010 | 2.510 |
| | (0.061) | (0.049) | (0.019) | (0.019) | (1.953) | (2.622) | (3.722) | (2.693) |
| Control | YES | YES | YES | YES | YES | YES | YES | YES |
| County FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations in pre-trends test regression | 2,802 | 2,865 | 3,923 | 3,828 | 2,236 | 2,086 | 1,667 | 1,486 |
| F-stat for all horizons | 0.17 | 1.35 | 2.25 | 2.59 | 8.36 | 0.88 | 0.05 | 0.47 |
| p(F) | 0.85 | 0.36 | 0.22 | 0.19 | 0.04 | 0.48 | 0.95 | 0.65 |

**Note:** Data from CHFS 2011-2019. Additional controls: urban, gender, education level, number of children in preschool, number of children in primary school, number of children in middle school, county per capita GDP, family size, whether living with grandparents, whether work in public sector, school canteen dummy. Year horizon 3 is the base year. Standard errors clustered by county and wave in parentheses. $* = p < 0.1$ ; $** = p < 0.05$; $*** = p < 0.01$.

# Chapter 3

# Absences from school and educational outcomes. Religious observance among ethnic minority students

**Angus Holford**[*] **Ziyi Huang**[†] **Birgitta Rabe**[‡]

**Abstract:** *This paper estimates the causal effect of school absences on the educational attainment of religiously observant ethnic minority students in England. We exploit exogenous variation in whether major Muslim and Hindu festivals fall on school days, generating absences unrelated to underlying achievement. Using administrative data and a two-stage least squares design, we show that festival-induced absences have no detectable effects in primary school but significantly reduce secondary students' English attainment and the probability of achieving five 'good' GCSEs. The estimated impacts are considerably larger than those found for general student populations, highlighting substantial heterogeneity in how absences affect learning.*

**Keywords:**Absence; Educational outcomes; Religious festival, Inequality, Educational policy
**JEL code:** I20, I24; I26, I28

[*]University of Essex and IZA.

[†]University of Essex. Corresponding author: Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ. zh17565@essex.ac.uk

[‡]Holford and Rabe acknowledge additional support from the ESRC Research Centre on Micro-Social Change [ES/S0124861], and exemplary research assistance from Hester Burn to compile the auxiliary datasets. This work uses data from the Department for Education's National Pupil Database, carried out in the Secure Research Service, part of the Office for National Statistics (ONS). 

## 3.1 Introduction

School attendance is widely viewed as fundamental to students' academic success. Consequently, absenteeism has long been a central policy concern, and most OECD countries operate a range of interventions—such as early warning systems, mental health support, and attendance advisors—to reduce school absences (OECD, 2024). Although a strong positive correlation between attendance and attainment is well established (e.g., Department for Education, 2025), credible causal evidence remains limited, as identification issues are challenging. Even less is known about how the consequences of absenteeism vary across student groups. It may be that catching up on lost learning during a day's absence is more difficult for some groups than others, but evidence on this is lacking.

This paper examines the causal effect of school absences on the educational attainment of ethnic minority students in England. Our identification strategy exploits exogenous variation in whether important Muslim and Hindu festivals fall on a school day, causing students to be absent from school. These festivals rotate into and out of school days according to lunar calendars and are typically not accounted for in the English school calendar. Focusing on religious festivals allows us to study a group that may struggle more than others to catch up on lost learning: religiously observant ethnic minority students. Compared with both the White British majority and non-observant ethnic minority students, they are more likely to be eligible for Free School Meals, live in disadvantaged areas, and speak a language other than English at home. In England, around one-third of ethnic minority students miss school at least once to observe a religious festival.

Absences may affect achievement through two channels. First, they directly reduce instructional time, which has been shown to affect educational outcomes, e.g. in studies on shortening the school week (Lavy, 2015; Thompson & Ward, 2022). Second, the fact that absences are not coordinated with other students is likely to affect students over and above the impact of losing instructional time (Goodman, 2014). Teachers will be constrained in helping students to catch up on missed lesson content once they return from an absence, while continuing with regular teaching of all students. The catch-up burden falls largely on the absent student, and incomplete catch-up may hinder subsequent learning. If catching up is more challenging for e.g. under-performing students or those with less resources and support at home, absences can potentially widen attainment gaps (Aucejo & Romano, 2016).

Credibly estimating the causal effect of absences on educational outcomes is challenging because absenteeism correlates with unobserved determinants of achievement (e.g. motivation, family shocks) and may itself be affected by prior performance. One strand of the literature has used fixed effects strategies to remove confounding factors at the school, class, teacher, household, or student-level (Aucejo & Romano, 2016; Cattan et al., 2023; Gershenson et al., 2017; Gottfried, 2009, 2011), and using within-student between-subject variation in absences to eliminate time-varying individual-level unobserved shocks (Liu et al., 2021). This literature finds detrimental contemporary effects on educational performance in the range of 2-5% of a standard deviation for ten days of absence in primary school (Aucejo & Romano, 2016; Cattan et al., 2023), and of similar magnitude in secondary schools in California (Liu et al., 2021).[4] Evidence on heterogeneity is mixed: some studies find larger negative effects for low- than high-income students (Aucejo & Romano, 2016; Gershenson et al., 2017), while others find no difference by background (Cattan et al., 2023; Liu et al., 2021).

A second strand of the literature - all for the US - uses instrumental variables exploiting exogenous shocks to attendance, such as influenza outbreaks (Aucejo & Romano, 2016), snowfall (Goodman, 2014), or transportation frictions proxied by distance to school (Gottfried, 2010). These studies also document significant effects of absences on achievement.[5] However, in some

[4]Specifically, Aucejo and Romano (2016) find that ten days of absence reduce mathematics achievement by between 5.5% and 19.8% of a standard deviation, and English achievement by between 2.9% and 11.3% of a standard deviation across different specifications. In their preferred specification, the effects are 5.5% of a standard deviation in mathematics and 2.9% in English. Cattan et al. (2023) find ten days of absence will decrease students' average grade points by 4.5% standard deviation. Liu et al. (2021), in their preferred model, show that ten absences in a subject lead to a reduction of 4.2% of a standard deviation in mathematics and 4% of a standard deviation in English Language Arts. Note that they examine absences by class period rather than by day; however, when each subject is taught only once per day, the number of class-period absences coincides with the number of days absent from that subject.

[5]Aucejo and Romano (2016) show that, in their fixed-effects model, ten days of absence lead to a reduction of 5.5% of a standard deviation in mathematics and 2.9% of a standard deviation in English. In their instrumental variable specification, the negative effects increase to 18.2% of a standard deviation in English. Goodman (2014) show a larger effect, finding that ten days of absence reduces students' mathematics scores by 52% of a standard deviation. Gottfried (2010) find that a one standard deviation increase in the number of days a student is present leads to 0.39–0.45 standard deviation increase in GPA in the baseline model. In their descriptive statistics the standard deviation of days present is 17 and the standard deviation of GPA is 1.1, then ten days of absence correspond to a 23–27% standard-deviation decrease in GPA.

cases the exclusion restriction may be violated (if flu directly impacts on achievement, for example) and the resulting local average treatment effects pertain to shocks that are not amenable to policy intervention (Aucejo & Romano, 2016).

We propose a new source of variation to instrument absences from school: religious festivals of a students' ethnic group falling on a school day rather than the weekend or a holiday.[6] This type of absence has the advantage that it is unlikely to be driven by health conditions or other unpredicted shocks, but it may differ from standard absences, as it combines missed instructional time with potential celebration-related effects such as fatigue, and may therefore have limited external validity beyond this specific context. After Christianity, the two religions with the highest population shares in England are Islam and Hinduism. Both religions have major festivals, such as Eid al-Fitr, Eid al-Adha and Diwali, that are not accounted for in the English school calendar which is traditionally arranged around the majority Christian religion and has holidays at festivals such as Easter and Christmas. Both Muslim and Hindu festivals follow a lunar calendar. This means they do not fall on the same weekday each year, like Easter for example, but instead can fall on school days in some years and weekends or holidays in others. We show that when major festivals fall on school days, absences among affected ethnic minority students rise by approximately 0.3 days per festival.

We use data from the National Pupil Database, an administrative data set of children in state-funded schools in England that contains students' grades as they progress through school and the counts of school sessions (of which there are two each day) missed in each school term. We focus on educational outcomes at the end of primary school (in grade 6, at age 11) where we observe achievement in Math and Reading and at the end of secondary school (in grade 11, at age 16) where we observe results achieved in General Certificate of Secondary Education (GCSEs) in Math and English, as well as a summary achievement measure, attaining 5 'good' GCSEs. Because we do not observe students' religion directly, we follow the previous literature in focusing on ethnic groups with high adherence to Islam or Hinduism (Almond et al., 2015).

[6] Religious festival timing has previously been used by Iyer and Shrivastava (2018) who use Hindu festivals falling on a Friday as an instrument for religious riots and whether it further impacts election results in India, and by Montero and Yang (2022) who use official celebration dates for Catholic saints prescribed by the global Catholic Church to instrument actual celebrated dates to study the effect of celebrations on agricultural outcomes in Mexico.

Based on these data, we estimate a two-stage least squares (2SLS) model of absences in the last year of primary and secondary school, respectively, on educational outcomes in the same year with time and school fixed effects[7]. We instrument absence days in test years using the number of the two main religious festival days of the students' ethnic group falling on a school day in that year.

To assess the validity of the instrument, we examine whether local authorities adjust school calendars to accommodate minority religious festivals, whether families sort across areas in response to festival timing, and whether fasting prior to Eid al-Fitr or air pollution associated with Diwali might directly affect outcomes. We find no evidence of systematic calendar manipulation or parental sorting, and show that neither Ramadan fasting nor Diwali-related pollution meaningfully affects test performance. A remaining limitation is that schools may use discretionary teacher training days to offset festival timing, especially in areas with high concentrations of affected students, though we cannot observe this directly. However, the first stage is stable across high- and low-minority schools at least for older students, suggesting limited scope for such adjustments; to the extent they occur, they would attenuate the first stage and render our estimates conservative.

Our results indicate that absences have sizeable effects for religiously observant ethnic minority students in secondary but not primary school. In primary school, absence days have no detectable effect on English or Mathematics outcomes. At the end of secondary school, however, there is a detrimental effect on English outcomes and on the likelihood to obtain five or more 'good' GCSEs, but no effect on math. One additional absence day reduces English scores by approximately 3-4% of a standard deviation, and reduces by 1-1.3 percentage points the chance of achieving 5 or more 'good' GCSEs for pupils whose absences respond to the timing of religious festivals. The impact of absences on observant ethnic minority students in secondary schools is therefore considerably larger than the effects estimated in prior studies for general student populations.

[7] We focus on same-year absences both because absences closest to the test date have the largest negative effects on achievement (Gottfried & Kirksey, 2017) and because the quasi-random timing of religious festivals generates variation at the annual level. Aggregating exposure across multiple years substantially attenuates the extent of residual identifying variation in the instrument: students in consecutive cohorts within the same Local Authority will share the same term and festival dates for all but one of their years at school, so pooling exposure across years leaves little year-specific variation on which identification can rest.

Our paper contributes to the literature in several ways. First, we provide the first causal evidence on the effects of absences on educational outcomes in England and the first analysis focused specifically on religiously observant ethnic minority students. As noted, this group faces substantial disadvantages which may exacerbate the difficulty of catching up on lost learning. Our findings therefore speak directly to settings, common across majority Christian countries such as Australia, Canada, France, or Germany, where Muslim and Hindu populations form sizeable minorities and may face similar challenges.[8] Moreover, our estimates show substantially larger absence effects for this population than the effects documented for general student populations in prior studies. This suggests a high degree of heterogeneity in how absences affect achievement, underscoring the need for further research on the groups most at risk and the potential for targeted policy interventions, such as catch-up support, for students who face multiple barriers to learning.

Second, we introduce a new instrumental variable for school absences. A key advantage of this instrument relative to those previously used, such as influenza outbreaks, snowfall, or distance to school, is that its first-stage variation is tied to a margin policymakers can actually influence. School calendars are adjustable, and several US school districts have recently incorporated minority-religion festivals into their schedules,[9] yet no evidence exists on the consequences of such adjustments. Our findings therefore speak to education authorities considering whether and how to recognise minority religious festivals in their local calendars.

Finally, our paper contributes to the literature on how religious observance interacts with schooling. Existing studies focus primarily on the contemporaneous effects of Ramadan fasting on test scores and exam performance (e.g. Oosterbeek & van der Klaauw, 2013; Hornung et al., 2023; Hanemaaijer et al., 2023), with only one paper examining how Ramadan affects school absences (Andersen & Houmark, 2025), and none linking such absences to academic achievement. Our focus on religiously observant students whose festival-related absences are shaped by the lunar calendar provides new evidence on how religious practices translate into differen-

[8]Muslim populations are 3.2% in Australia, 4.9% in Canada, 10% in France, and 6.6% in Germany, and Hindu populations 2.7% in Australia and 2.3% in Canada.(U.S. Department of State, 2024)

[9]E.g., Miami-Dade County School recognized Eid al-Fitr in 2023–24, Palm Beach County included Eid al-Fitr in 2024 and 2025, and similar actions have been taken in counties in New Jersey, Maryland, and Ohio (Molina, 2023; Sawchuk, 2020).

tial learning opportunities. More generally, our findings contribute to broader debates about the role of education institutions in shaping ethnic minority integration and achievement. Existing work has examined how school assignment rules, admission mechanisms, and education tracking affect minority and majority student outcomes (e.g. Guryan, 2004; Angrist & Lang, 2004; Brinkmann et al., 2024), as well as how curriculum content (e.g. Dee & Penner, 2017) and teacher bias or stereotyping (e.g. Alesina et al., 2024; Burn et al., 2024) influence academic performance. We introduce a novel institutional margin—term-date and holiday scheduling—that has received little attention despite its potential to create systematic differences in instructional time across groups.

The paper proceeds as follows. The next section describes the background of our study. Section 3 describes the data used, Section 4 lays out our method and probes identifying assumptions, and section 5 presents results and robustness checks. Section 6 concludes.

## 3.2 Background

### 3.2.1 Schooling in England

Schooling in England is compulsory from age 5 to 18, spanning primary school ("Reception" year plus grades 1-6), secondary school (grades 7-11) and a number of education routes post-16. Learning across primary and secondary school is divided into four Key Stages. At the end of primary school, at age 10/11, pupils in all government-funded primary schools participate in Key Stage 2 (KS2) National Curriculum assessments, also known as Standard Assessment Tasks (SATs), in mathematics, reading, writing, and science. At the end of secondary school, at age 16, pupils take Key Stage 4 (KS4) qualifications, mainly General Certificates of Secondary Education (GCSEs), in core subjects including English and math, and a range of optional subjects. To enable us to compare impacts of absences between education stages, we focus our analysis on exam results in English (reading) and math (see section 3.3.1).

The school year in state schools in England consists of 190 teaching days, or equivalently 380 morning or afternoon sessions. The school year runs from early September to mid-July, typically with two-week holidays at Christmas and Easter, one-week half-term breaks beginning in late October, February, and May, and an approximately 6 week summer holiday. In addition to holidays, there are five teacher training days each year during which schools are closed to pupils.

How the exact timing of school holidays is determined depends on the type of school. Local authorities set holiday dates for the schools they run, while academies and free schools, which are funded directly by the government and operate independently of local authorities, have the freedom to choose their calendars at the school level (Long, 2023). In practice, they mostly follow local authority calendars because many parents prefer that their children in different schools have the same holiday schedule, and teachers want their holidays to coincide with those of their own children. We therefore assume in our empirical analysis that academies and free schools follow the same holiday dates as local authority-run schools.[10] Moreover, state-funded schools in England designated as faith schools (75% of these being Church of England schools, see Long et al., 2024) can set their school calendar based on religious observances (Department for Education, 2024), and we exclude these from our analysis.

All schools are free to schedule teacher training days on dates that suit them, typically at the beginning of school terms. Schools could, however, schedule training days to coincide with major non-Christian festivals to avoid absences by ethnic minority pupils and teachers. This would likely be more prevalent in high ethnic minority-share areas. We do not observe the timing of teacher training days in our data, but we will present evidence that religious festivals increase student absences when they fall on a school day, and that there is little difference between areas with low and high shares of ethnic minority students in primary school and no difference in secondary school.

### 3.2.2 Hindu and Muslim festivals

Muslim festivals are based on the Hijri calendar, a lunar calendar with 354 or 355 days each year. Compared to the solar calendar, the Hijri calendar shifts forwards by approximately ten days every solar year. The two most significant holidays in Islam, celebrated by Muslims worldwide, are Eid al-Fitr and Eid al-Adha. Eid al-Fitr marks the end of Ramadan, the first

[10] We will show that results are robust to including academy-year FE in our estimations, which control for potential differences between academy and non-academy schools in the same year.

day on which Muslims break their fast (Islamic Relief, 2024b). Eid al-Adha is celebrated to remember the sacrifice made by Prophet Ibrahim and during which Muslims distribute meat among family, relatives, friends, and poor people (Islamic Relief, 2024a). Further festivals of relevance to Muslims include Hijri New Year, Mawlid al-Nabi, Ashura and Lailat al-Qadr.

The Hindu calendar, also known as the Panchanga, is a lunisolar calendar that determines the dates of Hindu festivals. Diwali is the most important and widely celebrated Hindu festivals that typically falls in October or November and lasts five days. The first two days are dedicated to preparations, and the third day is the main day of Diwali (Main Diwali), with celebrations occurring in the evening. The fourth day, Govardhan Puja, marks the New Year's Day for Hindus, and Hindus traditionally meet family and friends on this and the following day (Dabba, 2022). Students observing Diwali would therefore be most likely to be absent from school on the two days following the night-time celebrations of Main Diwali. Hindus also celebrate other festivals including Holi, Durga Puja, and Janmashtami.

To choose which Muslim and Hindu festivals to include in our analysis, we will examine empirically the relevance of the candidate festivals for student absences in England and include those that are most relevant (see Section 3.3.2).

## 3.3 Data

### 3.3.1 Data and outcomes

We use the National Pupil Database (NPD), an administrative dataset of state schools in England, covering 93% of students (about 600,000 students per cohort).[11] Data are reported for every student for each year that they are educated in schools, and contain students' educational outcomes as they progress through education, as well as absence data (collected since the 2006/2007 academic year) and student background characteristics.

Absences are classified as authorised or unauthorised, and within the authorised category absences are coded according to the reason of absence, including illness, medical appointments, family holidays and, importantly for our purposes, religious or cultural observance. There is a

[11] In England, 7% of students are educated in private schools, which are not government-funded and not required to submit data for inclusion in the National Pupil Database.

possibility that parents do not correctly report the reason for absence, for example opting to report an absence as illness when it is indeed a not agreed family holiday. We will therefore present results using both absence for religious or cultural observance and total authorised absences, the summary category. While absences are recorded in number of sessions (2 per day), we measure absences in days to align with festival days.

To the NPD we link a dataset of the school term and holiday dates set by local authorities in England since academic year 2006/07. We scrape this information from Local Authority websites using the Internet Archive maintained by Wayback Machine, and where necessary (for example due to changes in website structures or local government reorganizations) complement the information using Freedom of Information requests to current local authorities for information on term dates set by their predecessors. We then collect Muslim and Hindu festival dates using the website Time and Date (Time and Date AS, n.d.) and integrate these with our school term database for each academic year.

Academic outcomes in grade 6 (Key Stage 2) include students' test scores in mathematics and reading. To make test scores comparable across years, all test scores are standardised within each respective academic year with a mean zero and standard deviation of one. In grade 11 (Key Stage 4), we do not directly observe test scores in English and math. Instead, we observe grades ranging from A* to G, and from academic year 2016/17 from 9 to 1. Because these grades are not comparable over time, we end our observation window in 2015/16. We follow standard practice and convert A* to G grades into point scores which again we standardise separately by academic year to have a mean of zero and standard deviation of one.[12] In addition, in grade 11 we use a binary outcome often used to measure academic success, which codes as one all students who have passed at least five GCSEs or equivalent qualifications, including English and Math, and zero otherwise, where a pass grade is C or higher. Together, this is often taken as a measure of achieving 5+ 'good' GCSEs.

Our control variables include age in months to control for age at testing; a free school meal eligibility indicator, as eligibility may affect a student's motivation to attend school and serves as an indicator of a low-income background; the Income Deprivation Affecting Children Index

[12]Specifically, following Strand et al. (2015), GCSE test grades are transferred into scores by "U/X = 0, G = 16, F = 22, E = 28, D = 34, C = 40, B = 46, A = 52, A* = 58

(IDACI), defined as the proportion of children under 16 living in low-income households within each local area, which proxies children's socioeconomic background. We also include controls for students' gender, first language, and whether they need special educational support.

### 3.3.2 Sample

We restrict our analysis to the time-period from academic year 2006/07, when absences were first recorded in the NPD, to academic year 2018/19 (the last complete pre-pandemic year in the data) when considering grade 6 (Key Stage 2) outcomes, and to 2015/16 when considering grade 11 (Key Stage 4) outcomes as comparable outcome measures are not available after 2015/16. In each academic year, we also restrict the sample to students taking their Key Stage 2 and 4 exams in that year.

Next, we select the ethnic minority groups for whom Muslim or Hindu festivals may be relevant. We do not observe students' religion in the data, so we focus on ethnic groups with high adherence to Islam or Hinduism. According to the 2011 Census, Islam and Hinduism are the two largest Religions after Christianity in England, and 91% of Pakistani, 90% of Bangladeshi, and 51% of Other ethnic group are Muslim, while 44% of Indians are Hindus (Office for National Statistics, 2011).[13] No other ethnic group we can identify in the NPD has similarly high religious affiliation with Islam or Hinduism,[14] and we therefore restrict our sample to these four groups. They represent around 41% of the ethnic minority student population, and roughly 10% of the whole student population in England in 2018/19.

We assign Muslim festivals to students of Pakistani, Bangladeshi and Other ethnicity, and Hindu festivals to students of Indian ethnicity. To choose which festivals to include we take an empirical approach as follows. From student absence statistics presented in the next subsection we know that students of our ethnic groups of focus miss on average less than one day per academic year for religious observance. We therefore limit the candidate days to explore to the three most important festivals in each religion. We then regress different combinations of the three major religious festivals falling on school days on recorded absences for religious

[13]The Other ethnic groups in the ONS and the NPD differ slightly. In the ONS, the Other category includes individuals recorded as Roma, whereas these are listed as separate groups in the NPD.

[14]Mixed ethnicity, Chinese, other Asian and Black ethnicity populations all have Christianity as their most prevalent religion (Office for National Statistics, 2011).

observance with school and year fixed effects, separately for majority-Muslim and majority-Hindu groups. Table 3.1 shows the results of this exercise. The top panel shows that the fourth day of Diwali, "Govardhan," and the fifth day of Diwali, "Bhai," are the festivals most relevant to Hindu students' absences for religious observance.[15] The bottom panel shows the highest correlation between Muslim festivals and absences when the festivals are Eid al-Fitr and Eid al-Adha, and this correlation is slightly higher than when considering three festivals. Based on this, we proceed our analysis with the two most relevant religious festivals for each group.

**Table 3.1:** Association of religious festivals on absence for religious reasons

| | Hindu festivals | | | |
|---|---|---|---|---|
| | (1)<br>Diwali + Govardhan | (2)<br>Govardhan + Bhai | (3)<br>Diwali + Bhai | (4)<br>All three days |
| Festivals | 0.034***<br>(0.002) | 0.035***<br>(0.002) | 0.028***<br>(0.003) | 0.023***<br>(0.002) |
| School FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| N | 284,341 | 284,341 | 284,341 | 284,341 |
| | **Muslim festivals** | | | |
| | (1)<br>Eid al-Fitr + Eid al-Adha | (2)<br>Eid al-Fitr + Lailat | (3)<br>Eid al-Adha + Lailat | (4)<br>All three days |
| Festivals | 0.409***<br>(0.005) | 0.179***<br>(0.016) | 0.406***<br>(0.005) | 0.401***<br>(0.005) |
| School FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| N | 724,685 | 724,685 | 724,685 | 724,685 |

**Note:**Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Coefficients mean number of school days will be absent with religious reasons if festivals are falling on school days. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

To derive our final estimation sample, we clean the data of outliers, excluding students whose total absences exceed 120 sessions in a given academic year, which amounts to missing nearly an entire term (0.16% of students in grade 6 and 2.26% in grade 11). We also exclude students who do not meet the appropriate age criteria, specifically those who are not 10 years old at the beginning of grade 6 and not 15 years old at the beginning of grade 11 (0.15% of students

[15]Diwali is a multi-day festival, with the main religious celebration taking place in the evening of the third day, which is the main Diwali and is labelled as "Diwali" in the table. As a result, absences are more likely to occur on the fourth and fifth days, when families engage in New Year celebrations and social visits.

in grade 6 and 0.68% in grade 11 fall into this category). Finally, we exclude students attending faith schools (0.43% of students in grade 6 and 1.10% in grade 11 attend the 24 Muslim and 5 Hindu state-funded schools) because their school calendars are flexible and likely to close during festivals of their faith group to allow for religious observance.

### 3.3.3 Descriptive statistics

Figure 3.1 shows in Panel A the number of absence days per student in grade 6 and 11, respectively, over our observation period. Absences are considerably higher among grade 11 than grade 6 students. The number of absence days has reduced for both groups over time, with the most marked decline taking place between 2006/07 and 2013/14, possibly as a result of policies to combat absenteeism implemented in this time-period (Wallace, 2025). Panel B of the Figure shows the average number of absence days for religious or cultural observance. While there is a slight downward trend in these absences, clearly they follow a specific pattern - presumably the timing of religious festivals.

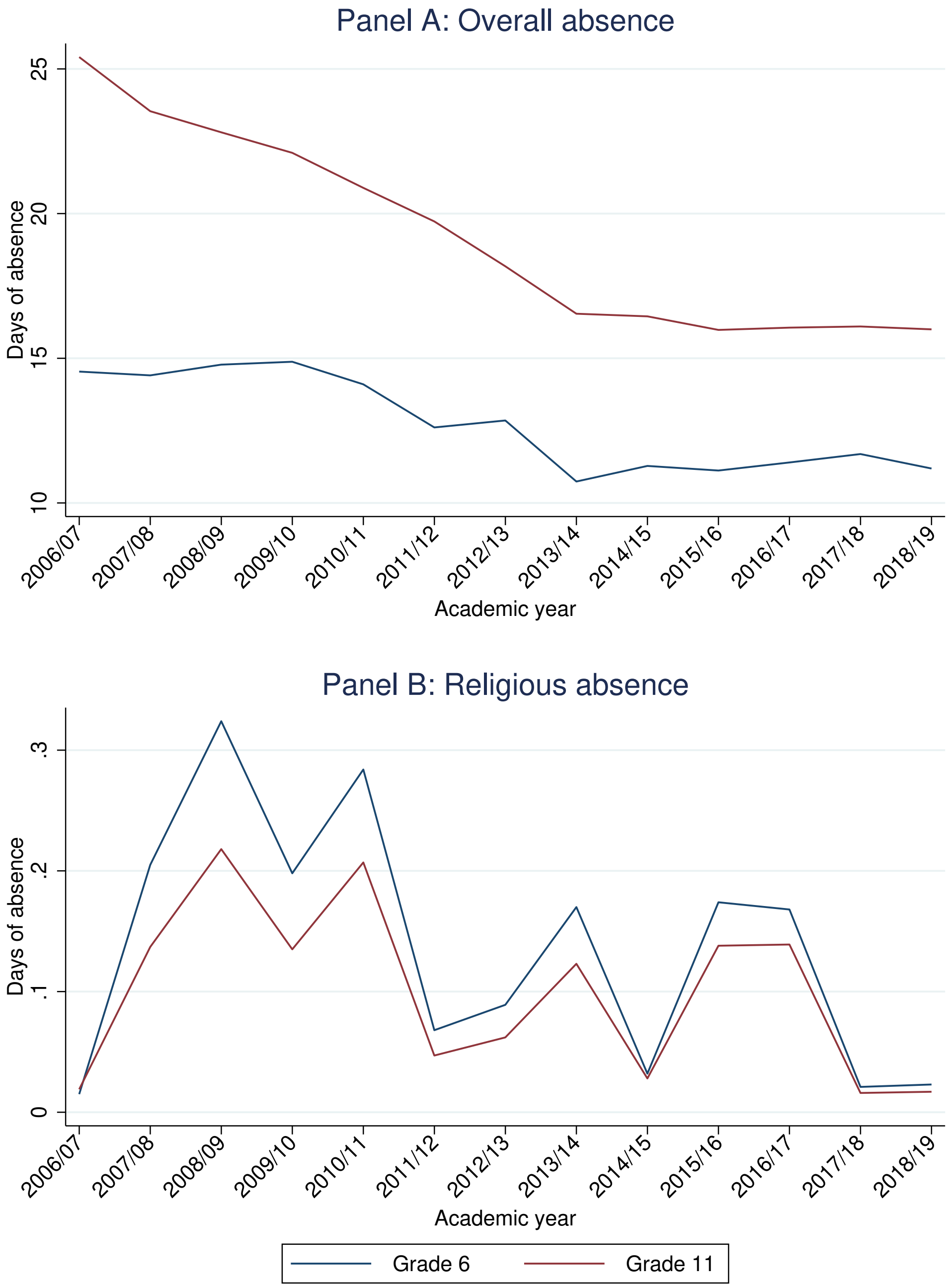
This is explored in Figure 3.2, which plots the number of main festivals (zero, one or two) falling on school days in a particular academic year alongside the absences for religious observance among our ethnic groups with high Muslim affiliation (Panel A) and high Hindu affiliation (Panel B). The figure shows that absences for religious observance among students from majority Muslim ethnic groups closely follow the pattern of festivals falling on school days, with students in grade 6 (primary school) slightly more likely to be absent for this reason than students in grade 11 (secondary school).[16] These patterns are also evident for Indian students (the majority Hindu group) in Panel B, though absences are considerably less aligned – perhaps expected, given that a lower proportion of Indians (44%) practice Hinduism according to the national statistics cited above. As noted, parents report the reason for any absence of their children, leaving room for misreporting of the reason. Appendix Table A3.1 breaks down absences by reason for absence for the whole student population and separately for our four ethnic groups of focus, showing a higher frequency of absences for religious reasons among our ethnic groups of focus than among students overall.

[16]The number of festivals falling on school days are not identical for all students because school calendars vary locally and across years. The numbers shown are averaged across affected students.

Table 3.2 presents characteristics and outcomes for students observed in grade 6. The equivalent Table for grade 11 is shown in Appendix Table A3.2 and displays similar patterns as in grade 6. Column (1) shows means for the whole population of students, column (2) shows means for our ethnic groups of focus, and column (3) the difference between the two groups. Along most background characteristics, our ethnic groups of focus are more disadvantaged than the student population overall: they are less likely to speak English as their first language, more likely to be eligible for free school meals and to live in a disadvantaged neighborhoods, but less likely to have special educational needs. The bottom two rows of the Table show that in grade 6 our ethnic groups of focus perform less well in reading than the overall student population but outperform the whole population in math. Appendix Table A3.2 shows equivalent patterns in educational outcomes in grade 11, with our ethnic minority groups performing slightly below average in English, but above average in Math and in the summary achievement measure (5+ 'good' GCSEs).

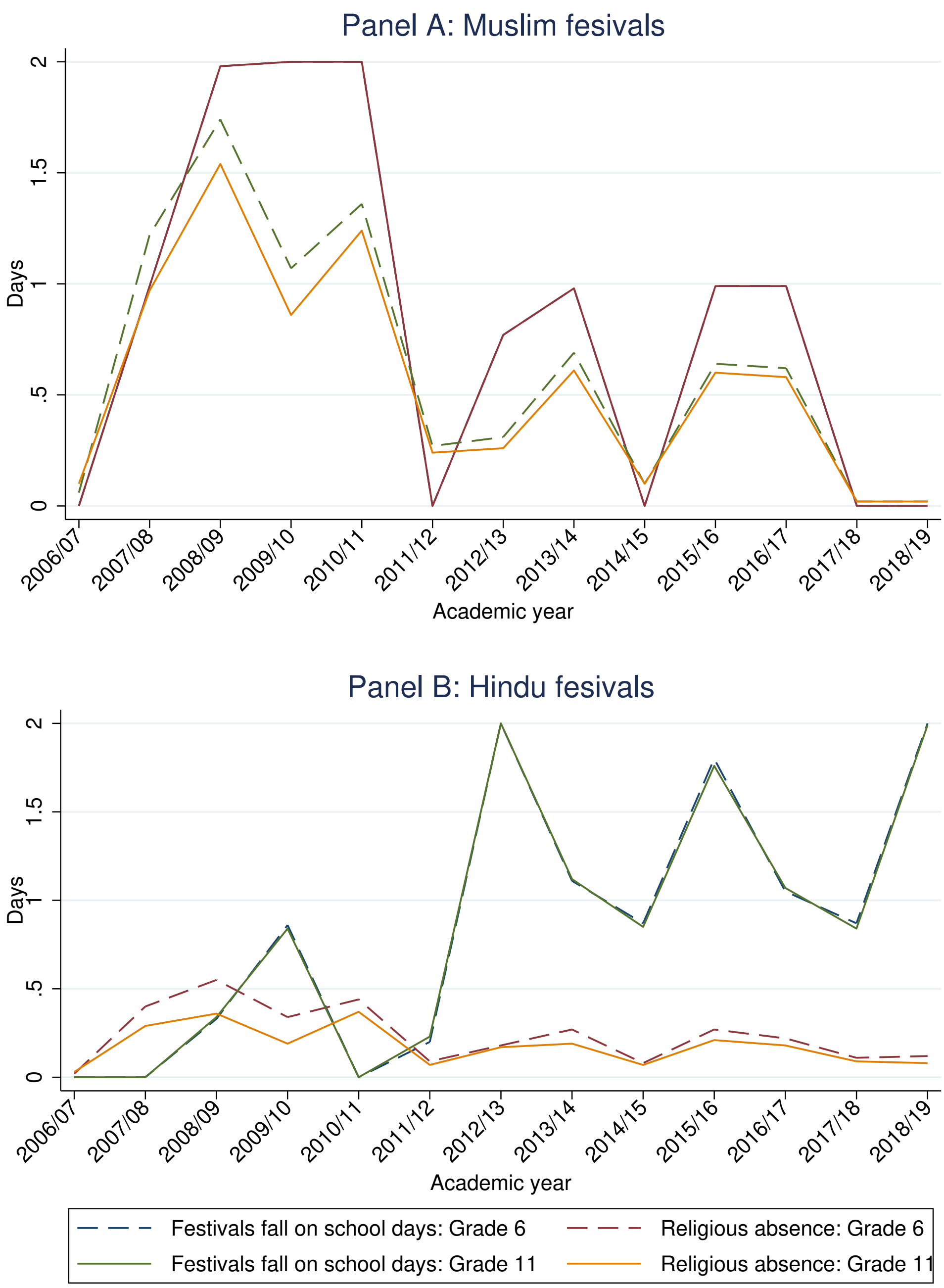
Table 3.2 further splits our ethnic groups of focus into a group of students that never missed school for religious or cultural observance (column 4) and a group of students that has (column 5). The group that has ever missed school for religious observance is presumably at least to some extent religiously observant and therefore will approximately characterise the group of compliers in our analysis. Differences between the groups are displayed in column (6) and show that religiously observant ethnic minority students (as proxied by having missed school for religious observation) are more disadvantaged on all background characteristics we can observe than non-observant ethnic minority students. They also have worse educational outcomes in English and math (and in the corresponding outcomes in grade 11). This shows that our analysis focuses on a group of particular policy interest, that may find it more difficult to catch up on lost learning than other groups.

**Figure 3.1:** School absences over time




**Note:** Data from NPD 2007–2019 for grade 6; Data from NPD 2007–2016 for grade 11.

**Figure 3.2:** Religious festivals and absences for religious observance




**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absences shown are for majority-Muslim ethnic minority groups (Bangladeshi and Pakistani and Other ethnicity) in panel A and majority-Hindu ethnic group (Indian) in Panel B.

**Table 3.2:** Descriptive statistics: Grade 6

| | Whole population | | | Within ethnic groups of focus | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Whole population | Ethnic groups of focus | Difference | Never missed school for religious or cultural observance | Ever miss school for religious or cultural observance | Difference |
| Gender (male = 1) | 0.511 | 0.513 | 0.002*** | 0.512 | 0.513 | 0.001 |
| | (0.500) | (0.500) | | (0.500) | (0.500) | |
| First language (English = 1) | 0.830 | 0.145 | -0.685*** | 0.223 | 0.114 | -0.109*** |
| | (0.376) | (0.352) | | (0.416) | (0.318) | |
| Free school meal (eligible = 1) | 0.167 | 0.209 | 0.042*** | 0.130 | 0.239 | 0.109*** |
| | (0.373) | (0.406) | | (0.336) | (0.427) | |
| SEN (need = 1) | 0.215 | 0.186 | -0.030*** | 0.151 | 0.199 | 0.049*** |
| | (0.411) | (0.389) | | (0.358) | (0.399) | |
| IDACI score | 0.222 | 0.324 | 0.102*** | 0.269 | 0.345 | 0.076*** |
| | (0.171) | (0.175) | | (0.174) | (0.172) | |
| Reading score | 0.000 | (0.128) | -0.128*** | 0.039 | -0.193 | -0.232*** |
| | (1.000) | (0.997) | | (0.997) | (0.989) | |
| Math score | 0.000 | 0.050 | 0.050*** | 0.252 | -0.028 | -0.281*** |
| | (1.000) | (1.005) | | (0.962) | (1.010) | |
| N | 7,442,481 | 728,530 | | 204,302 | 524,228 | |

**Note:** Data from NPD 2007-2019. Sample of students in grades 6. Whole population is comparing pupils from minor ethnic group and the whole population. Within ethnic groups of focus is comparing pupils who ever missed for religious reasons and those who never missed. Number of observations is the max observations where different percent of missing is existing in different controls. Scores are standardized by minus mean and divided by standard deviation.

## 3.4 Empirical strategy

We estimate the effect of student absences on educational outcomes using the following linear two-way fixed effects (TWFE) regression:

$$\text{Educ}_{tsi} = \text{Absence}_{tsi}\beta_1 + X_{tsi}\beta_2 + \tau_t + \mu_s + u_{tsi}, \tag{1}$$

where $\text{Educ}_{tsi}$ are academic outcomes including Math and Reading scores at KS2 and GCSE outcomes at KS4, and $\text{Absence}_{tsi}$ is the number of absence days of student $i$ in year $t$ at school $s$, which we will measure using both total days absent and days absent for religious observance. $X_{tsi}$ are pupil- and household-level characteristics including age in months, eligibility for free school meals, local deprivation index, first language English, and special educational needs. $\tau_t$ is a year fixed effect and $\mu_s$ is a school fixed effect. $u_{tsi}$ is the error term.

The set of demographic characteristics controls for individual and household-level factors that may affect both absences and outcomes. For example, children from deprived neighbourhoods may find it both harder to travel to school and to access the learning resources to do well in school. The year fixed effects eliminate factors such as national absence policies that vary over time but not across schools and students, such as national absence policies, and the school fixed effects eliminate any time-invariant school-level influences, such as a rigorous enforcement of rules that may affect both absences and educational outcomes.

Even with this rich set of observable characteristics and fixed effects, the regression of absence days on academic performance may be biased due to omitted variables. For instance, poor underlying health may increase absenteeism and also reduce ability to learn, or parents prepared to take their children out of school for religious observance may be highly motivated and invest otherwise in their children's education. To address such endogeneity issues, we employ an instrumental variable (IV) strategy, where we instrument a student's number of days absent using the number of the two main Muslim or Hindu festivals (Eid al-Fitr and Eid al-Adha for Muslims and Govardhan Puja and Bhai Duj for Hindus) falling on a school day. Assuming these lunar-based festivals are not taken into account when determining school dates, the number of festival days that coincide with the school term will be plausibly random.

The first stage of our two-stage least squares (2SLS) estimation relates festival dates to absences as follows:

$$\text{Absence}_{tsi} = \text{Festival}_{ts}\,\beta_1 + X_{tsi}\,\beta_2 + \tau_t + \mu_s + \epsilon_{tsi}, \tag{2}$$

where $\text{Absence}_{tsi}$ is again the number of absence days for student $i$ in year $t$ at school $s$. $\text{Festival}_{ts}$ is the number of festival days falling on school days, with $\text{Festival}_{ts} \in \{0, 1, 2\}$, and the other variables are as before. $\epsilon_{tsi}$ is the error term.

The regression for the second stage is:

$$\text{Educ}_{tsi} = \widehat{\text{Absence}}_{tsi}\beta_1 + X_{tsi}\beta_2 + \tau_t + \mu_s + \upsilon_{tsi}, \tag{3}$$

where $\widehat{\text{Absence}}_{tsi}$ is the predicted level of absence from the first stage. Our coefficient of interest is $\beta_1$, the effect of an additional day of absence on educational outcomes. Under the standard IV assumptions, the 2SLS estimate of $\beta_1$ can be interpreted as the local average treatment effect (LATE) for students whose absence behaviour is shifted by the instrument. This estimate captures a net effect, as absences may improve the performance of remaining students through smaller class sizes and increased teacher attention, thereby lowering an absent student's relative performance; but they may also reduce peer quality and limit opportunities for classroom interaction and peer learning, which would tend to offset this effect.

After applying year and school fixed effects in our regression, the identifying variation comes from differences in the alignment between religious holidays and school calendars across years within the same school. Note that because schools are nested within local authorities, school fixed effects fully absorb local authority level fixed effects (recall that the instrument does not vary within LAs in a given year). Table 3.3 shows the identifying variation in the instrument in terms of means and standard deviations of the variable $\text{Festival}_{ts}$ (a religious festival falling on a school day), separately by Muslim and Hindu festivals. In grade 6, shown in the top panel, an average of 0.7 main Muslim festivals, out of a maximum of two, falls on a school day across years, with a standard deviation of 0.74. For Hindu festivals the average is 0.97 days with a standard deviation of 0.80. After adding year fixed-effects, the standard deviation of Muslim festivals falling on school days decreases to 0.14, and for Hindu festivals to 0.38. Adding school fixed-effects and the demographic control variables included in our 2SLS estimation hardly changes the standard deviation. These results suggest that we still have variation in the IV after controlling for year fixed-effects, school fixed-effects, and controlling for demographic characteristics. The bottom panel shows equivalent descriptives for students in the year of their Key Stage 4 exams.

**Table 3.3:** Identifying variation: residual standard deviation of festivals falling on school days

| Grade 6 | Muslim: Eid al-Fitr and Eid al-Adha | | | Hindu: Govardhan Puja and Bhai Dui | | |
|---|---|---|---|---|---|---|
| | Mean | Standard deviation | N | Mean | Standard deviation | N |
| Raw | 0.735 | 0.736 | 472,399 | 0.974 | 0.796 | 174,389 |
| + year FE | 0 | 0.143 | 472,399 | 0 | 0.376 | 174,389 |
| + School FE | 0 | 0.144 | 472,399 | 0 | 0.376 | 174,389 |
| + other controls | 0 | 0.144 | 471,239 | 0 | 0.376 | 174,077 |

| Grade 11 | Muslim: Eid al-Fitr and Eid al-Adha | | | Hindu: Govardhan Puja and Bhai Dui | | |
|---|---|---|---|---|---|---|
| | Mean | Standard deviation | N | Mean | Standard deviation | N |
| Raw | 0.932 | 0.758 | 254,720 | 0.801 | 0.831 | 114,313 |
| + year FE | 0 | 0.166 | 254,720 | 0 | 0.430 | 114,313 |
| + School FE | 0 | 0.166 | 254,720 | 0 | 0.430 | 114,313 |
| + other controls | 0 | 0.166 | 254,138 | 0 | 0.430 | 114,145 |

**Note:**Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Variation for Hindu festivals only includes Indian student, and Variation for Muslim festivals includes Bangladeshi, Pakistani, and Other ethnicity. Other controls include age in month; eligibility for school meals; IDACI index; gender, first language; whether need special education.

## Threats to identification

The number of religious festivals falling on school days must satisfy two conditions to be a valid instrument: (i) independence — festival timing must be unrelated to unobserved determinants of student achievement; and (ii) the exclusion restriction — festival timing must affect outcomes only through absence. We discuss threats to each condition in turn.

A first concern is that local authorities (LAs) may adjust school calendars in response to the ethnic composition of their student population. If LAs aim to accommodate religious observance —for example for groups whose educational outcomes are of concern — they could set term dates to reduce the number of festival days falling on school days. This would imply a negative selection of students into festival-related absences and induce correlation between the instrument and the error term. To examine this, we regress the number of Muslim and Hindu festival days falling on school days by local authority and year on LA-level shares of majority Muslim or majority Hindu ethnic groups, under progressively rich specifications. Across specifications in Table 3.4, we find no evidence that school term dates respond to the religious composition of the student body.

A related concern is that individual schools may use their discretionary INSET days, of which there are five per year, to offset the occurrence of religious festivals on school days. Such adjustments would be more likely in schools with higher concentrations of students from the relevant religious groups, and would weaken the link between festival timing and actual absence. Because the timing of INSET days is not observed in our data, we cannot directly test this mechanism. However, we show later that the first-stage relationship between festival timing and pupil absence does not differ significantly between schools with high and low densities of affected ethnic-minority students. Nonetheless, we cannot completely rule out such behaviour. If it were occurring, it would tend to attenuate the first stage—reducing variation in absence induced by festival timing—and therefore bias the IV estimates toward zero. Our estimates should therefore be interpreted as conservative with respect to the true effect of festival-related absence on educational outcomes.

Another potential threat is parental sorting. Parents who prefer alignment between school holidays and religious festivals may (i) attempt to influence LA decisions on term dates, or (ii) relocate to LAs where school dates already align more closely with religious observance. The first channel would imply stronger influence in areas with more members of the same religious group; however, as shown above, school calendars do not appear responsive to demographic variation. To examine the second channel, we test whether future festival timing predicts changes in the ethnic composition of LAs. Table 3.5 shows results of regressing changes in shares of relevant ethnic groups in local authorities on religious festivals falling on school days with year and local authority fixed effect in each of the next six school years in column (1) for majority Muslim groups and column (3) for majority Hindu groups. Columns (2) and (4) show results when summing the number of festivals on school days over the next six academic years.

Across the two specifications for each ethnic group, we find no evidence of systematic differential out-migration related to festival timing. If anything, we find some evidence of in-migration rising when festivals fall on school days in selected years, but these relationships seem spurious and too small in magnitude to pose a threat to identification.

A final threat concerns possible direct effects of festival timing on educational outcomes, violating the exclusion restriction. Two channels warrant consideration. First, the Diwali period is associated with increased air pollution due to fireworks, which could affect students' short-term health and therefore their academic performance (Asif et al., 2024; Buwaniwal et al., 2023). However, Hindu festival dates fall largely in October–November, well outside the May testing period, and any pollution effects are not tied to whether the festival falls on a school day, so should occur independently of festival timing. We later show in robustness checks that there is no evidence that additional Hindu festival days falling on school days increase illness-related absences, suggesting that this channel does not induce a direct effect on outcomes.

Second, Ramadan has been shown to affect concentration and fatigue, and thereby learning and test performance (Hanemaaijer et al., 2023; Hornung et al., 2023; Oosterbeek & van der Klaauw, 2013). Over the study period, Ramadan occurred between June and January, and key testing dates for Key Stage 2 and 4 exams fell before Ramadan in most years. Moreover, if fasting were to influence learning over a longer period, beyond its effect on test performance, this mechanism would operate regardless of whether the Eid al-Fitr festival, which marks the end of Ramadan fasting, falls on a school day or not, and therefore would not be correlated with the instrument. In robustness checks we present estimation results excluding the years in which testing overlapped with Ramadan that are qualitatively similar to our main findings.

**Table 3.4:** Check for validity of the IV: effect of share of ethnic on festival falling on school days

| | Eid al-Fitr and Eid al-Adha fall on school day | | | Govardhan Puja and Bhai Dui fall on school day | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| | OLS | Year FE | Year and LA FE | OLS | Year FE | Year and LA FE |
| Indian students' share | | | | 0.005 | 0.001 | -0.017 |
| | | | | (0.004) | (0.001) | (0.017) |
| Bangladeshi and Pakistani and other ethnic students' share | -0.004 | 0.001 | 0.001 | | | |
| | (0.003) | (0.001) | (0.006) | | | |
| Year FE | NO | YES | YES | NO | YES | YES |
| LA FE | NO | NO | YES | NO | NO | YES |
| N | 1,739 | 1,739 | 1,739 | 1,739 | 1,739 | 1,739 |
| Number of year | | 13 | 13 | | 13 | 13 |
| Number of LA | | | 152 | | | 152 |

**Note:** Data from NPD 2007-2019. All samples are pooled. The coefficient means how many festival days will fall on school days if 1% of specific ethnicity share increase in the local authority. Standard errors in parentheses, and clustered at the LA level. * = ”p<0.1”, ** = ”p<0.05”, *** = ”p <0.01”.

**Table 3.5:** Check for validity of the IV: Effect of religious festival on changing of Ethnic Groups of Focus students' share

| | (1)<br>Change of Muslim share | (2)<br>Change of Muslim share | (3)<br>Change of Hindu share | (4)<br>Change of Hindu share |
|---|---|---|---|---|
| Sum of all forward Religious festivals | | 0.104 | | 0.025* |
| | | (0.064) | | (0.013) |
| Religious festivals: 6 year forward | 0.086 | | 0.024 | |
| | (0.152) | | (0.023) | |
| Religious festivals: 5 year forward | 0.187* | | 0.03 | |
| | (0.101) | | (0.023) | |
| Religious festivals: 4 year forward | -0.035 | | 0.058*** | |
| | (0.097) | | (0.021) | |
| Religious festivals: 3 year forward | 0.083 | | 0.007 | |
| | (0.086) | | (0.021) | |
| Religious festivals: 2 year forward | 0.140* | | 0.005 | |
| | (0.081) | | (0.026) | |
| Religious festivals: 1 year forward | 0.091 | | 0.013 | |
| | (0.084) | | (0.023) | |
| Year FE | YES | YES | YES | YES |
| LA FE | YES | YES | YES | YES |
| N | 799 | 799 | 799 | 799 |
| Number of year | 6 | 6 | 6 | 6 |
| Number of LA | 152 | 152 | 152 | 152 |

**Note:** Data from NPD 2007-2019. All samples are pooled. The coefficient means how many percentage of Hindu/Muslim will move if religious festivals fall on school days after X years. "Change of share" is calculated by using current percentage minus percentage in the last year. "Sum of all forward Religious festivals" is a aggregated number of religious festivals falling on school days in next 6 years. "Hindu share" is the share of religious observant Indian ethnicity students, and "Muslim share" is the share of religious observant Muslim ethnicity students. Standard errors in parentheses, and clustered at the LA level. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

## 3.5 Results

### 3.5.1 Religious festivals and absences from school

We begin by exploring the first-stage relationship between festivals falling on school days and absences from school among students who are from majority-Muslim or majority-Hindu ethnic groups. Table 3.6 shows the contemporaneous effects on overall absences and on absences for religious observance in grade 6 in columns (1) and (2), and the equivalent effects in grade 11 in columns (3) and (4). Results in column (1) for grade 6 show that for each additional religious festival day that coincides with a school day, students in the sample miss 0.274 days of school. The coefficient in column (2) for absences for religious observance is identical to that in column (1) for overall absences, suggesting that parents correctly report the reason for absence being religious observance at this age. For grade 11, a higher point estimate in column (3) indicates that religious festivals increase overall absences by slightly more at age 15/16 than at age 10/11 (in column 1), though this difference is not statistically significant. The slightly lower coefficient in column (4) suggests that in grade 11 some parents may record festival-related absences as happening for other reasons than for religious observance, for example as an illness or family holiday. This confirms that both absence classifications should be used in our analysis. The F-statistics for both measures of absences and in both grades 6 and 11 are significantly greater than 10, indicating that they are strong and the variation we exploit is relevant.

### 3.5.2 Absences and education outcomes

Table 3.7 and Table 3.8 present the effects of absences driven by religious festivals on the educational outcomes of pupils in majority-Muslim and majority-Hindu ethnic groups in grades 6 and 11, respectively. In each table, we report both OLS and IV estimates, with the upper panels focusing on total absences and the lower panels on absences for religious observance.

For grade 6, the fixed-effects OLS estimates in columns (1) and (2) of the top panel of Table 3.7 show negative and statistically significant associations between absences and both reading and mathematics performance. Once we account for endogeneity using our IV strategy,

**Table 3.6:** First stage results: Effects of religious festivals falling on school day on absences

| | Grade 6 | | Grade 11 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Overall absence | Absence for religious observance | Overall absence | Absence for religious observance |
| Religious festivals | 0.274*** | 0.274*** | 0.330*** | 0.249*** |
| | (0.019) | (0.009) | (0.042) | (0.015) |
| Control | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| School FE | Yes | Yes | Yes | Yes |
| F-stats for religious festivals | 219.37 | 865.69 | 60.72 | 294.39 |
| P(F) | 0.00 | 0.00 | 0.00 | 0.00 |
| N | 602,840 | 602,840 | 354,073 | 354,073 |
| Number of schools | 11,374 | 11,374 | 3,396 | 3,396 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absence counts in days. Controls include age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. Standard errors in parentheses, and clustered at the school level. * = ”p<0.1”, ** = ”p<0.05”, *** = ”p <0.01”.

however, these relationships disappear: the IV coefficients are close to zero and imprecisely estimated. The larger negative OLS coefficients for mathematics suggest that pupils with weaker underlying performance are more likely to be absent, and this is confirmed by the endogeneity test, which rejects exogeneity of absences for mathematics. The lower panel of Table 3.7 reveals a similar pattern when focusing on absences for religious observance. Taken together, the grade 6 results imply that short, isolated absences of this type do not measurably affect pupils' contemporaneous performance in primary school.

Table 3.8 reports the corresponding results for grade 11. Here, OLS estimates in the upper panel show strong negative associations between absences and outcomes in mathematics, English, and the composite indicator of achieving at least five 'good' GCSEs. After instrumenting absences, the mathematics effect is no longer present, but sizeable and statistically significant negative effects remain for English and for the probability of securing five or more 'good' GCSEs. Results are similar when we focus specifically on absences for religious observance in the bottom panel of the Table.

Across the two panels of Table 3.8, our results show that each additional day of absence from school reduces standardised English scores by around 3–4% of a standard deviation. While previous typically report effects of 2–5% of a standard deviation for ten days of absence (Aucejo & Romano, 2016; Cattan et al., 2023; Liu et al., 2021), the comparison is not straightforward.

Our estimates refer to the impact of one day of absence, but we cannot simply scale them up to compare with effects for ten days because doing so would extrapolate beyond the 0–2-day range that identifies our estimates. Nonetheless, the findings indicate that even short absences have larger negative effects on English performance than those found for whole populations in previous papers.

As shown in Table 3.1 and Appendix Table A3.2, the ethnic groups we study perform above the national average in mathematics but below average in reading (grade 6) and English (grade 11). It may therefore be easier for these pupils to catch up in mathematics than in English, which has greater linguistic demands and is not the first language spoken at home for around 80% of our sample. This difficulty may also help explain the broader impact on attainment: missing one day of school reduces the probability of achieving five or more ‘good’ GCSEs by 1–1.3 percentage points.

The contrast between grades 6 and 11 suggests that the consequences of missing school intensify as pupils progress into a more structured and assessment-driven phase of schooling. By the final year of secondary education, the curriculum builds more tightly on prior lessons, and the proximity of high-stakes exams leaves little room for re-teaching or consolidation. Under these conditions, even small disruptions can hinder progress, especially in language-intensive subjects.

**Table 3.7:** Effect of absences on pupils' educational outcomes: Grade 6

| | OLS | | IV | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Math | Reading | Math | Reading |
| **Overall absence** | | | | |
| Days of absence | -0.015*** | -0.007*** | 0.011 | -0.007 |
| | (0.000) | (0.000) | (0.012) | (0.011) |
| Controls | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| School FE | YES | YES | YES | YES |
| F-stats for religious festivals | | | 219.37 | 219.37 |
| Endogeneity test F-stats | | | 5.16 | 0.00 |
| P(F) | | | 0.02 | 0.98 |
| **Absence for religious observance** | | | | |
| Days of absence | -0.014*** | -0.009*** | 0.011 | -0.007 |
| | (0.003) | (0.003) | (0.012) | (0.011) |
| Controls | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| School FE | YES | YES | YES | YES |
| F-stats for religious festivals | | | 865.69 | 865.69 |
| Endogeneity test F-stats | | | 5.29 | 0.03 |
| P(F) | | | 0.02 | 0.87 |
| N | 602,840 | 602,840 | 602,840 | 602,840 |
| Number of schools | 11,374 | 11,374 | 11,374 | 11,374 |

**Note:** Data from NPD 2007-2019. Sample of students in grades 6. Absence counts in days. Controls include age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. Outcomes are from SATs, and standardized within year by minus mean and divided by standard deviation. The endogeneity test is the Wu-Hausmann test. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

**Table 3.8:** Effect of absences on pupils' educational outcomes: Grade 11

| | OLS | | | IV | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | 5+ 'good' GCSEs | Math | English | 5+ 'good' GCSEs | Math | English |
| **Overall absence** | | | | | | |
| Days of absence | -0.011*** | -0.029*** | -0.022*** | -0.010* | -0.006 | -0.031*** |
| | (0.000) | (0.000) | (0.000) | (0.005) | (0.011) | (0.012) |
| Control | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| School FE | Yes | Yes | Yes | Yes | Yes | Yes |
| F-stats for Religious festivals | | | | 60.72 | 60.72 | 60.72 |
| Endogeneity test F-stats | | | | 0.12 | 5.11 | 0.6 |
| P(F) | | | | 0.72 | 0.02 | 0.44 |
| **Absence for religious observance** | | | | | | |
| Days of absence | -0.022*** | -0.056*** | -0.046*** | -0.013* | -0.008 | -0.042** |
| | (0.002) | (0.004) | (0.004) | (0.007) | (0.014) | (0.016) |
| Control | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES |
| School FE | YES | YES | YES | YES | YES | YES |
| F-stats for Religious festivals | | | | 294.39 | 294.39 | 294.39 |
| Endogeneity test F-stats | | | | 1.72 | 11.23 | 0.07 |
| P(F) | | | | 0.19 | 0.00 | 0.79 |
| N | 354,073 | 354,073 | 354,073 | 354,073 | 354,073 | 354,073 |
| Number of schools | 3,396 | 3,396 | 3,396 | 3,396 | 3,396 | 3,396 |

**Note:** Data from NPD 2007-2016. Sample of students in grades 11. Absence count in days. Controls includes age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. GCSE test grades are transferred into scores by "U/X = 0, G = 16, F = 22, E = 28, D = 34, C = 40, B = 46, A = 52, A* = 58". Scores are standardized within in year by minus mean and divided by standard deviation. The endogeneity test is the Wu-Hausmann test. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

### 3.5.3 Heterogeneous effects

To learn more about how absences for religious observance may differently affect different groups we estimate heterogenous effects by interacting our treatment with binary variables relating to their major religion (Islam/Hinduism), gender, eligibility for free school meals, first language, and whether they study in a school with a high or low share of Hindu or Muslim students (above/below median share). Table 3.9 presents the first-stage results. They show that in primary school (grade 6) students from majority-Muslim ethnic groups are considerably more likely to miss school when festivals fall on school days than students from the majority-Hindu ethnic group, but this difference is no longer apparent in secondary school, in grade 11. Table 3.9 also shows differences in primary school by gender (with girls slightly more likely to be absent for religious observance than boys), by ethnic share (with students in low-minority share schools more likely to miss school than in high-minority share schools), and by first language (with those not speaking English at home more likely to miss school than those speaking English). None of these differences are apparent in secondary school.

Note that the higher first-stage coefficient for students in schools with a low (compared to high) share of students from ethnic minority groups of focus could indicate that high-share schools may accommodate religious festivals of their student population through timing of the teacher INSET days. This is indicated in primary school, while in secondary school the difference is smaller and no longer statistically significant. As noted in the introduction, this attenuation of the first stage would bias our estimates downwards.

The top panel of Table 3.9 shows a low F statistic for the first stage by Religion group. We therefore omit this distinction in our 2SLS IV estimates. These are displayed in Table 3.10. Across the distinction by gender, free school meal eligibility, ethnic shares and first language spoken at home, we find statistically significant group differences for most second-stage effects in primary school and some in secondary school. This points to different effects for the compliers in each group.

In primary school (grade 6), boys, children eligible for free school meals, in high ethnic-minority-share schools and whose first language is English are more negatively affected by a festival-induced absences than their counterparts. We also observe some positive effects for girls and for students in low ethnic-minority-share schools. This may reflect differences in compensatory behaviour. In schools with a low share of ethnic minority students, absences for religious and cultural observance will be less common, so parents may be more conscious that their child is losing a day of school relative to their peers. This is more likely to trigger parental responses, such as additional study effort at home, which can offset the loss of instructional time. In addition, girls' high levels of consciousness (Carvalho, 2016) may make them able to compensate for missed learning.

In secondary school (grade 11), differences are less often statistically significant and vary by outcome measure. Focusing on English, where we saw statistically significant detrimental effects across the students in our sample, we again find that boys and free school meal eligible children fare worse, though the differences are not statistically significant. Students in high ethnic-minority-share schools and whose first language is not English are significantly more negatively affected than low ethnic-minority-share and English speaking students. In sum, while no entirely clear pattern emerges across grades and outcomes, results suggest that students from lower-income families where English is not spoken as the first language and attending schools with high shares of ethnic minority students have a harder time catching up on lost learning than other students.

**Table 3.9:** Heterogeneity in first-stage effects of instrument.

| | Overall absence | |
|---|---|---|
| | Grade 6 | Grade 11 |
| Muslim | 0.409*** | 0.338*** |
| | (0.030) | (0.056) |
| Hindu | 0.157*** | 0.324** |
| | (0.024) | (0.061) |
| Difference | -0.253*** | -0.014 |
| | (0.039) | (0.082) |
| Weak IV test F-stat | 8.96 | 0.42 |
| Females | 0.294*** | 0.317*** |
| | (0.021) | (0.049) |
| Males | 0.255*** | 0.343*** |
| | (0.021) | (0.047) |
| Difference | -0.040* | 0.026 |
| | (0.020) | (0.046) |
| Weak IV test F-stat | 252.52 | 61.92 |
| No FSM eligibility | 0.279*** | 0.343*** |
| | (0.018) | (0.042) |
| FSM eligibility | 0.251*** | 0.281*** |
| | (0.034) | (0.065) |
| Difference | -0.028 | -0.062 |
| | (0.031) | (0.053) |
| Weak IV test F-stat | 238.36 | 81.21 |
| Low ethnic minority share | 0.315*** | 0.365*** |
| | (0.020) | (0.044) |
| High ethnic minority share | 0.231*** | 0.298*** |
| | (0.026) | (0.061) |
| Difference | -0.084*** | -0.067 |
| | (0.028) | (0.066) |
| Weak IV test F-stat | 275.17 | 89.00 |
| Other first language | 0.288*** | 0.319*** |
| | (0.020) | (0.046) |
| English first language | 0.208*** | 0.364*** |
| | (0.029) | (0.055) |
| Difference | -0.080*** | 0.045 |
| | (0.030) | (0.055) |
| Weak IV test F-stat | 63.08 | 67.11 |
| N | 602,840 | 354,073 |
| Number of schools | 11,374 | 3,396 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absence count in days. Coefficients are from regression on the pooled sample with interaction term; festivals and festivals*dummy are the instruments for absence and absence*dummy. This is the first stage for absence. Controls includes age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. The weak IV test is Sanderson–Windmeijer test which is conducted for each endogenous regressor in models with multiple endogenous variables. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** = "p <0.01".

**Table 3.10:** Treatment heterogeneity

| | Overall absence | | | | |
|---|---|---|---|---|---|
| | Grade 6 | | Grade 11 | | |
| | Math | Reading | 5+ ’good’ GCSEs | Math | English |
| Females | 0.021* | -0.008 | -0.011* | -0.013 | -0.028** |
| | (0.012) | (0.011) | (0.006) | (0.011) | (0.013) |
| Males | 0.002 | -0.006 | -0.008 | 0.001 | -0.034** |
| | (0.013) | (0.012) | (0.006) | (0.011) | (0.012) |
| Difference | -0.019*** | 0.002 | 0.002 | 0.014* | -0.006 |
| | (0.007) | (0.006) | (0.004) | (0.008) | (0.008) |
| Heteroskedasticity-robust test F-stat | 108.57 | 108.57 | 30.41 | 30.41 | 30.41 |
| No FSM eligibility | 0.017 | -0.001 | -0.011** | -0.013 | -0.029** |
| | (0.012) | (0.011) | (0.005) | (0.010) | (0.012) |
| FSM eligibility | 0.005 | -0.014 | -0.008 | -0.001 | -0.033** |
| | (0.013) | (0.012) | (0.006) | (0.012) | (0.013) |
| Difference | -0.012** | -0.013** | 0.003 | 0.012* | -0.004 |
| | (0.006) | (0.006) | (0.003) | (0.007) | (0.007) |
| Heteroskedasticity-robust test F-stat | 111.54 | 111.54 | 32.09 | 32.09 | 32.09 |
| Low ethnic minority share | 0.023** | 0.012 | -0.005 | -0.010 | -0.009 |
| | (0.010) | (0.010) | (0.006) | (0.011) | (0.012) |
| High ethnic minority share | 0.005 | -0.017 | -0.012** | -0.004 | -0.042*** |
| | (0.014) | (0.014) | (0.006) | (0.012) | (0.014) |
| Difference | -0.018 | -0.029*** | -0.007 | 0.006 | -0.033** |
| | (0.012) | (0.011) | (0.005) | (0.012) | (0.013) |
| Heteroskedasticity-robust test F-stat | 121.99 | 121.99 | 37.85 | 37.85 | 37.85 |
| Other first language | -0.008 | -0.025* | -0.010* | -0.006 | -0.032*** |
| | (0.014) | (0.014) | (0.005) | (0.011) | (0.012) |
| English first language | -0.165*** | -0.170*** | -0.007 | -0.011 | -0.007 |
| | (0.058) | (0.060) | (0.007) | (0.014) | (0.016) |
| Difference | -0.157*** | -0.145*** | 0.003 | -0.005 | 0.025* |
| | (0.049) | (0.050) | (0.006) | (0.012) | (0.013) |
| Heteroskedasticity-robust test F-stat | 6.84 | 6.84 | 26.94 | 26.94 | 26.94 |
| N | 602,840 | 602,840 | 354,073 | 354,073 | 354,073 |
| Number of schools | 11,374 | 11,374 | 3,396 | 3,396 | 3,396 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absence count in days. Coefficients are from regression on the pooled sample with interaction term; festivals and festivals*dummy are the instruments for absence and absence*dummy. Controls includes age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. GCSE test grades are transferred into scores by “U/X = 0, G = 16, F = 22, E = 28, D = 34, C = 40, B = 46, A = 52, A* = 58”. Scores are standardized within year by minus mean and divided by standard deviation. The heteroskedasticity-robust test is Kleibergen-Paap Wald test of weak instruments for the overall identification strength of the IV. Standard errors in parentheses, and clustered at the school level. * = ”p<0.1”, ** = ”p<0.05”, *** = ”p <0.01”.

## 3.6 Robustness checks

### Other ethnic groups

We conduct a check to examine whether religious festivals affect the absence of all pupils, rather than focusing just on majority-Muslim and majority-Hindu ethnic groups. Here we include all ethnicities in the estimation sample, and construct an instrument comprising the count across the four Hindu and Muslim festivals falling on school days. The results are reported in Table 3.11. Reassuringly, these festivals do not have a significant effect on absences for all pupils. This rules out the timing of religious festivals on school days being coincident with other unobserved factors affecting absence more generally, or that school or classroom-based responses to the festivals have a effect on every students.

**Table 3.11:** Impact of Hindu and Muslim Religious Festivals on overall pupil absences

| | Overall absence | |
|---|---|---|
| | Grade 6 | Grade 11 |
| **All ethnicities** | | |
| Hindu plus Muslim Religious festivals | -0.003 | 0.014 |
| | (0.012) | (0.053) |
| F-stats for Religious festivals | 0.05 | 0.06 |
| P(F) | 0.82 | 0.8 |
| N | 6,310,682 | 4,639,480 |
| Number of schools | 17,804 | 4,370 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absence counts in days. Controls include age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education. In Academy-Year FE, year fixed-effect replaces by academy-year fixed-effect. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** ="p <0.01".

### Ramadan fasting and air pollution from Diwali celebrations

We discussed in section 3.4 that there may be independent effects of Muslim and Hindu festivals on educational outcomes through the effects of Ramadan fasting and of Diwali air pollution-induced illness, invalidating the exclusion restriction. While, as noted, these effects would be expected to affect students regardless of whether the festivals fall on school days or not in a particular year, we perform the following tests. First we examine whether the number of Hindu festivals falling on school days increases students' absence due to illness or medical reasons, consistent with the air pollution hypothesis. Results in Table 3.12 show no significant festival impact on student absences for illness or medical reasons, suggesting that any increase in air pollution associated with Diwali celebrations does not negatively affect students' health, or if it does, it is not directly related to the festival falling on a school day.

**Table 3.12:** Illness and medical absence

| | Grade 6 | | Grade 11 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Illness absence | Medical absence | Illness absence | Medical absence |
| Hindu religious festivals | 0.014 | 0.003 | 0.016 | -0.008* |
| | (0.011) | (0.002) | (0.021) | (0.004) |
| Control | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| School FE | YES | YES | YES | YES |
| F-stats for religious festivals | 1.42 | 2.62 | 0.55 | 2.97 |
| P(F) | 0.23 | 0.11 | 0.46 | 0.09 |
| N | 602,840 | 602,840 | 354,073 | 354,073 |
| Number of schools | 11,374 | 11,374 | 3,396 | 3,396 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Absence counts in days. Controls include age in month; eligibility for school meal; IDACI index; gender; first language; whether need special education. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** ="p <0.01".

Second, we investigate whether Ramadan fasting may affect our results. The Ramadan period moves forward every year, and first started overlapping with exam dates from academic year 2015/16, as shown in Table 3.13. 2015/16 is the last year of our observation window for Key Stage 4 (grade 11) exams. In the years 2017/18 and 2018/19 Ramadan overlaps with Key Stage 2 (grade 6) exams.

**Table 3.13:** Ramadan date and end of tests

| Year | SATs end date | GCSE end date | Ramadan start date | Grade 6's and grade 11's tests affected by Ramadan |
|---|---|---|---|---|
| 2014 | 15/05/2014 | 23/06/2014 | 29/06/2014 | None |
| 2015 | 14/05/2015 | 17/06/2015 | 18/06/2015 | None |
| 2016 | 12/05/2016 | 22/06/2016 | 07/06/2016 | Grade 11 (GCSE) |
| 2017 | 11/05/2017 | | 27/05/2017 | None |
| 2018 | 17/05/2018 | | 17/05/2018 | Grade 6 (SATs) |
| 2019 | 16/05/2019 | | 06/05/2019 | Grade 6 (SATs) |

**Note:** The date for the SATs will be no later than May, and the date for the GCSEs will be no later than June. The Ramadan start date will move about 10 days early each year, and it falls after June from 2007-2013.

To check whether the Ramadan fasting period overlapping with exam dates affects outcomes, we estimate our 2SLS instrumental variable regression while omitting the affected years.

Table 3.14 shows that results are very similar when dropping the affected years, suggesting that Ramadan fasting does not affect results differently in years that festivals fall on school days rather than weekend or the school holidays.

**Measurement of school holidays**

We next check potential error in our measure of school holidays and term-time. Schools can set 5 teacher training dates that may be used flexibly across the academic year – something we cannot observe in our data. Training days are often placed at the beginning of each school term, so we drop the first day of each term in our analysis to check whether results remain similar to our baseline estimates.

A further concern is that academy schools have the autonomy to set their own school date, so have more freedom to accommodate religious festivals of minority groups which may weaken the relationship between religious festivals and absences. To control for this, we replace the year fixed effects in our 2SLS estimation with academy-year fixed effects, where 'academy' is a binary variable indicating whether a school is an academy or local authority-run school.

Table 3.14 shows that the effect of absence on students' academic outcomes remains similar to the baseline results when dropping training days and when using academy-year fixed effects, although the academy-year fixed effects slightly reduce the point estimates for grade 11. This suggests that academies may adopt measures that reduce the negative effects of festival-related absences, but the overall relationships remain close to those identified in the main analysis.

**Alternative grade 11 sample**

For estimates relating to grade 11 outcomes we use a shorter observation window than for grade 6 as there was a grade reform introduced in 2016/17 for Key Stage 4 tests that are not comparable pre reform grades. New grades range from 9 to 1, while old grades range from A* to G. To test whether results remain similar when using the whole available time-period, we convert old grades to new numerical scores following Department for Education (2019), coding them as: U/X = 0, G = 1, F = 1.5, E = 2, D = 3, C = 4, B = 5.5, A = 6, and A* = 8.5. Results in Table 3.14 shows that no meaningful difference in the first- or second-stage results.

**Table 3.14:** Robustness check

| | Grade 6 | | | Grade 11 | | | |
|---|---|---|---|---|---|---|---|
| | First stage: Overall absence | Second stage: Math | Reading | First stage: Overall absence | Second stage: 5+ 'good' GCSEs | Math | English |
| **Baseline results** | | | | | | | |
| Religious festivals | 0.274*** | | | 0.330*** | | | |
| | (0.019) | | | (0.042) | | | |
| Overall absence | | 0.011 | -0.007 | | -0.010* | -0.006 | -0.031*** |
| | | (0.012) | (0.011) | | (0.005) | (0.011) | (0.012) |
| F-stats in the first stage | 219.37 | | | 60.72 | | | |
| N | 602,840 | 602,840 | 602,840 | 354,073 | 354,073 | 354,073 | 354,073 |
| Number of schools | 11,374 | 11,374 | 11,374 | 3,396 | 3,396 | 3,396 | 3,396 |
| **Before Fasting** | | | | | | | |
| Religious festivals | 0.263*** | | | 0.353*** | | | |
| | (0.021) | | | (0.042) | | | |
| Overall absence | | 0.003 | 0.003 | | -0.010** | -0.012 | -0.034*** |
| | | (0.015) | (0.014) | | (0.005) | (0.010) | (0.012) |
| F-stats in the first stage | 152.07 | | | 69.68 | | | |
| N | 468,987 | 468,987 | 468,987 | 306,354 | 306,354 | 306,354 | 306,354 |
| Number of schools | 10,382 | 10,382 | 10,382 | 3,268 | 3,268 | 3,268 | 3,268 |
| **With training days** | | | | | | | |
| Religious festivals | 0.263*** | | | 0.322*** | | | |
| | (0.019) | | | (0.041) | | | |
| Overall absence | | 0.020 | -0.001 | | -0.010* | -0.008 | -0.034*** |
| | | (0.013) | (0.012) | | (0.005) | (0.011) | (0.012) |
| F-stats in the first stage | 201.83 | | | 60.69 | | | |
| N | 602,840 | 602,840 | 602,840 | 354,073 | 354,073 | 354,073 | 354,073 |
| Number of schools | 11,374 | 11,374 | 11,374 | 3,396 | 3,396 | 3,396 | 3,396 |
| **Academy-Year FE** | | | | | | | |
| Religious festivals | 0.273*** | | | 0.322*** | | | |
| | (0.018) | | | (0.041) | | | |
| Overall absence | | 0.011 | -0.008 | | -0.008 | 0.000 | -0.027** |
| | | (0.012) | (0.011) | | (0.005) | (0.011) | (0.012) |
| F-stats in the first stage | 219.37 | | | 60.76 | | | |
| N | 602,204 | 602,204 | 602,204 | 351,598 | 351,598 | 351,598 | 351,598 |
| Number of schools | 11,307 | 11,307 | 11,307 | 3,309 | 3,309 | 3,309 | 3,309 |
| **Alternative grade 11 sample** | | | | | | | |
| Religious festivals | | | | 0.310*** | | | |
| | | | | (0.036) | | | |
| Overall absence | | | | | -0.007 | 0.017 | -0.030*** |
| | | | | | (0.005) | (0.012) | (0.012) |
| F-stats in the first stage | | | | 74.39 | | | |
| N | | | | 512,634 | 512,634 | 512,634 | 512,634 |
| Number of schools | | | | 3,833 | 3,833 | 3,833 | 3,833 |

**Note:** In first panel: data from NPD 2007-2017, and sample of students in grade 6; data from NPD 2007-2015 and sample of students in grade 11. In second and third panel: data from NPD 2007-2019, and sample of students in grade 6; data from NPD 2007-2016 and sample of students in grade 11; In fourth panel: data from NPD 2007-2019, and sample of students in grade 6; data from NPD 2007-2019 and sample of students in grade 11. Absence count in days. Controls includes age in month; eligibility for school meals; IDACI index; gender; first language; whether need special education.School fixed effect and year fixed effect are included. In "Academy-Year FE", year FE is replaced by Academy-Year FE. GCSE test grades are transferred into scores in "Alternative grade 11 sample" by "U/X = 0, G = 1, F = 1.5, E = 2, D = 3, C = 4, B = 5.5, A = 6, A* = 8.5". Scores are standardized within year by minus mean and divided by standard deviation. Standard errors in parentheses, and clustered at the school level. * = "p<0.1", ** = "p<0.05", *** ="p <0.01".

## 3.7 Conclusion

This paper provides new causal evidence on the consequences of school absences for the educational attainment of religiously observant ethnic minority students in England. Exploiting quasi-random variation in whether major Muslim and Hindu festivals fall on school days, we document that festival timing generates meaningful increases in absences among affected students. These absences have no detectable impact on achievement in primary school, but in secondary school—where learning is more hierarchical and exam-focused—they significantly reduce English achievement and lower the probability of attaining five 'good' GCSEs. The magnitude of these effects is substantially larger than those estimated in previous studies for the general student population (Aucejo & Romano, 2016; Cattan et al., 2023; Liu et al., 2021), suggesting that disadvantaged, observant minority students face particular challenges in catching up on missed instruction.

Our study has several limitations. First, although we find no evidence of systematic school-calendar adjustments or parental sorting, we cannot fully rule out the possibility that discretionary teacher training days attenuate the first stage, making our primary-school estimates conservative. Second, our identification strategy yields local average treatment effects for students whose absences respond to festival timing; these effects may not generalise to all absences or all student groups. Third, while we rule out several potential direct impacts of festivals, such as fasting and air pollution, other unobserved mechanisms could influence outcomes, although any such effects would likely bias estimates toward zero.

Despite these limitations, our findings have clear implications for educational policy. One avenue is for local authorities to consider accommodating major non-Christian religious festivals within school calendars, as is already done in some school districts in the US, for example. Recognising these festivals could reduce absence spikes and help equalise effective instructional time across groups. A second approach is to strengthen catch-up support for students who do miss school on festival days. This might include structured post-absence tutoring, targeted in-class reinforcement, or short-term adjustments to lesson sequencing in high-stakes years. More broadly, schools serving diverse communities may benefit from formal protocols for managing out-of-sync absences, ensuring that pupils returning from religious observance are not disproportionately disadvantaged.

Ultimately, our results highlight that the timing of the school year, an often overlooked institutional design choice, can shape learning opportunities in ways that interact strongly with cultural and religious practices. As ethnic and religious diversity continues to grow across many countries, understanding and addressing these interactions will be essential for designing equitable education systems.

# References


Alesina, A., Carlana, M., La Ferrara, E., & Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, *114*(7), 1916–1948.

Almond, D., Mazumder, B., & Van Ewijk, R. (2015). In utero ramadan exposure and children's academic performance. *The Economic Journal*, *125*(589), 1501–1533.

Andersen, C., & Houmark, M. A. (2025). Ramadan, absence and educational performance. *Available at SSRN 5268800*.

Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? evidence from boston's metco program. *American Economic Review*, *94*(5), 1613–1634.

Asif, M., Bhatti, S. M., Dhuria, S. R., & Yadav, S. (2024). Source apportionment of metal ions in ambient air (pm2.5) during firecracker bursting: A case study of amritsar diwali on 24 october 2022. *Urban Climate*, *53*, 101796.

Aucejo, M. E., & Romano, F. T. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55*, 70–87.

Brinkmann, M., Teltemann, J., Huth-Stöckle, N., & Schunck, R. (2024). Does between-school tracking increase school segregation among migrant students? a difference-in-differences and multiverse analysis of international large-scale assessment data. *Compare: A Journal of Comparative and International Education*, *55*(6), 1106–1122.

Burn, H., Fumagalli, L., & Rabe, B. (2024). Stereotyping and ethnicity gaps in teacher assigned grades. *Labour Economics*, *89*, 102577.

Buwaniwal, A., et al. (2023). Size-segregated aerosol measurements during diwali festival in an elevated background location. *Atmospheric Environment*, *314*, 120078.

Carvalho, R. G. G. (2016). Gender differences in academic achievement: The mediating role of personality. *Personality and Individual Differences*, *94*, 54–58.

Cattan, S., Kamhofer, A. D., Karlsson, M., & Nilsson, T. (2023). The long-term effects of student absence: Evidence from sweden. *The Economic Journal*, *133*(650), 888–903.

Dabba, S. (2022, October). *Diwali*. https://studentlife.lincoln.ac.uk/2022/10/24/diwali/

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal*, *54*(1), 127–166.

Department for Education. (2019). *Secondary accountability measures: Guide for maintained secondary schools, academies and free schools*.

Department for Education. (2024, May). *How reducing the cap on faith school admissions will help to raise standards*.

Department for Education. (2025, March). *The link between attendance and attainment in an assessment year: Research report* (Research and analysis). Department for Education. https://dera.ioe.ac.uk/id/eprint/41140/

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are student absences worth the worry in us primary schools? *Education Finance and Policy*, *12*(2), 137–165.

Goodman, J. (2014). *Flaking out: Student absences and snow days as disruptions of instructional time* (tech. rep. No. w20221). National Bureau of Economic Research.

Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 392–415.

Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal*, *47*(2), 434–465.

Gottfried, M. A. (2011). The detrimental effects of missing school: Evidence from urban siblings. *American Journal of Education*, *117*(2), 147–182.

Gottfried, M. A., & Kirksey, J. J. (2017). “When” students miss school: The role of timing of absenteeism on students’ test performance. *Educational Researcher*, *46*(3), 119–130.

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, *94*(4), 919–943.

Hanemaaijer, K., Marie, O., & Musumeci, M. (2023). *The fast and the studios? ramadan observance and student performance* (tech. rep. No. 16249). IZA Discussion Paper.

Hornung, E., Schwerdt, G., & Strazzeri, M. (2023). Religious practice and student performance: Evidence from ramadan fasting. *Journal of Economic Behavior & Organization*, *205*, 100–119.

Islamic Relief. (2024a). *Eid al-adha / qurbani*. https://www.islamic-relief.org.uk/giving/islamic-giving/qurbani/eid-al-adha/

Islamic Relief. (2024b). *Eid al-fitr*. http://www.islamic-relief.org.uk/giving/islamic-giving/ramadan/eid-al-fitr/

Iyer, S., & Shrivastava, A. (2018). Religious riots and electoral politics in india. *Journal of Development Economics*, *131*, 104–122.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? evidence from developed and developing countries. *The Economic Journal*, *125*(588), F397–F424.

Liu, J., Lee, M., & Gershenson, S. (2021). The short- and long-run impacts of secondary school absences. *Journal of Public Economics*, *199*, 104441.

Long, R. (2023). *The school day and year* (No. 07148). House of Commons Library.

Long, R., Roberts, N., & Maisuria, A. (2024). *Faith schools: Faqs* (No. 006792). House of Commons Library.

Molina, A. (2023, March). *Across the country, a push to observe muslim holidays in school calendars*. https://religionnews.com/2023/03/17/across-the-country-a-push-to-observe-muslim-holidays-in-school-calendars/

Montero, E., & Yang, D. (2022). Religious festivals and economic development: Evidence from the timing of mexican saint day festivals. *American Economic Review*, *112*(10), 3176–3214.

OECD. (2024). Evaluating post-pandemic education policies and combatting student absenteeism beyond covid-19. *OECD Education Policy Perspectives*, (101). https://doi.org/10.1787/a38f74b2-en

Office for National Statistics. (2011). *Census 2011 – dc2201ew: Ethnic group by religion*. https://www.nomisweb.co.uk/census/2011/DC2201EW/view/2092957699?rows=c_ethpuk11&cols=c_relpuk11

Oosterbeek, H., & van der Klaauw, B. (2013). Ramadan, fasting and educational outcomes. *Economics of Education Review*, *34*, 219–226.

Sawchuk, S. (2020, March). *Schools reconsider the calendar as students grow more diverse*. https://www.edweek.org/leadership/schools-reconsider-the-calendar-as-students-grow-more-diverse/2020/03

Strand, S., Malmberg, L., & Hall, J. (2015). *English as an additional language and educational achievement in england: An analysis of the national pupil database* [Report]. University of Oxford.

Thompson, P. N., & Ward, J. (2022). Only a matter of time? the role of time in school on four-day school week achievement impacts. *Economics of Education Review*, *86*, 102198.

Time and Date AS. (n.d.). *Holidays today and upcoming holidays in the united kingdom*. https://www.timeanddate.com/holidays/uk

U.S. Department of State. (2024). *2023 report on international religious freedom* (tech. rep.). U.S. Department of State. https://www.state.gov/reports/2023-report-on-international-religious-freedom/

Wallace, M. (2025). *Reducing school absence: Innovation lessons from the last labour government*. Institute for Government.

# Appendix

**Table A3.1:** Proportions of absence for different reasons by ethnicity

| | Grade 6 | | | | | |
|---|---|---|---|---|---|---|
| | Whole population | Ethnic Groups of Focus: Bangladeshi | Indian | Pakistani | Other ethnicity | Total Ethnic Groups of Focus |
| Absence rate | 4.59% | 4.55% | 4.08% | 4.84% | 4.28% | 4.51% |
| of which: | | | | | | |
| Illness (Not medical or dental appointments) | 62.63% | 56.37% | 56.46% | 52.11% | 51.09% | 53.70% |
| Medical or dental appointments | 5.79% | 4.67% | 5.05% | 4.81% | 5.95% | 5.01% |
| Religious observance | 1.06% | 8.62% | 3.96% | 9.68% | 5.39% | 7.54% |
| Study leave | <0.1% | <0.1% | <0.1% | <0.1% | <0.1% | <0.1% |
| Agreed family holiday | 6.87% | 3.11% | 8.53% | 3.96% | 5.88% | 5.16% |
| Agreed extended family holiday | 0.14% | 0.66% | 0.75% | 0.74% | 0.47% | 0.69% |
| Family holiday, not agreed, or is taking days in excess of an agreed family holiday | 4.71% | 7.24% | 8.93% | 9.51% | 8.11% | 8.82% |
| Arrived after registers closed | 1.18% | 0.86% | 0.61% | 0.84% | 1.16% | 0.84% |
| Total authorized absence | 84.70% | 79.70% | 83.02% | 77.95% | 76.72% | 79.25% |
| Total unauthorized absence | 15.30% | 20.30% | 16.98% | 22.05% | 23.28% | 20.75% |
| Mean absent sessions | 12.7 | 12.76 | 11.06 | 13.78 | 11.74 | 12.57 |
| Total absent sessions | ¿94,600,000 | 1,564,066 | 2,130,674 | 4,122,019 | 1,339,282 | 9,156,041 |
| Observations | 7,442,481 | 122,591 | 192,620 | 299,228 | 114,091 | 728,530 |
| | Grade 11 | | | | | |
| | Whole population | Ethnic Groups of Focus: Bangladeshi | Indian | Pakistani | Other ethnicity | Total Ethnic Groups of Focus |
| Absence rate | 7.11% | 5.87% | 5.25% | 6.47% | 6.20% | 5.97% |
| of which: | | | | | | |
| Illness (Not medical or dental appointments) | 48.24% | 42.28% | 45.41% | 40.63% | 39.16% | 41.90% |
| Medical or dental appointments | 6.73% | 6.56% | 7.64% | 6.36% | 7.25% | 6.86% |
| Religious observance | 0.55% | 6.87% | 2.58% | 8.14% | 3.68% | 5.81% |
| Study leave | 9.71% | 5.06% | 15.59% | 6.40% | 7.91% | 8.81% |
| Agreed family holiday | 1.89% | 0.85% | 2.39% | 1.57% | 2.12% | 1.76% |
| Agreed extended family holiday | <0.1% | 0.10% | 0.11% | 0.19% | 0.11% | 0.14% |
| Family holiday, not agreed, or is taking days in excess of an agreed family holiday | 1.57% | 1.67% | 2.09% | 2.92% | 3.19% | 2.55% |
| Arrived after registers closed | 1.81% | 2.27% | 1.39% | 1.65% | 1.88% | 1.72% |
| Total authorized absence | 78.35% | 71.07% | 82.88% | 74.23% | 71.87% | 75.62% |
| Total unauthorized absence | 21.65% | 28.93% | 17.12% | 25.77% | 28.13% | 24.38% |
| Mean absent sessions | 20.23 | 16.64 | 14.57 | 18.78 | 17.36 | 16.95 |
| Total absent sessions | >113,000,000 | 1,156,504 | 1,928,832 | 3,187,235 | 1,148,400 | 7,420,971 |
| N | 5,569,170 | 69,485 | 132,409 | 169,705 | 66,148 | 437,747 |

**Note:** Data from NPD 2007-2019, and sample of students in grade 6; Data from NPD 2007-2016 and sample of students in grade 11. Percentages and means are calculated per student per year for all exam year from 2006/07 to 2018/19 for grade 6, and from 2006/07 to 2015/16 for grade 11. <0.1% includes all percentage smaller than 0.1%.

**Table A3.2:** Descriptive statistics: Grade 11

| | Whole population | | | Within ethnic groups of focus | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Whole population | Ethnic groups of focus | Difference | Never missed school for religious or cultural observance | Ever miss school for religious or cultural observance | Difference |
| Gender (male = 1) | 0.511 | 0.517 | 0.006*** | 0.517 | 0.517 | 0.000 |
| | (0.500) | (0.500) | | (0.500) | (0.500) | |
| First language (English = 1) | 0.878 | 0.214 | -0.665*** | 0.256 | 0.187 | -0.069*** |
| | (0.327) | (0.410) | | (0.437) | (0.390) | |
| Free school meal (eligible = 1) | 0.132 | 0.247 | 0.115*** | 0.181 | 0.288 | 0.107*** |
| | (0.339) | (0.432) | | (0.385) | (0.453) | |
| SEN (need = 1) | 0.201 | 0.194 | -0.008*** | 0.182 | 0.201 | 0.018*** |
| | (0.401) | (0.395) | | (0.386) | (0.400) | |
| IDACI score | 0.217 | 0.343 | 0.125*** | 0.296 | 0.371 | 0.075*** |
| | (0.175) | (0.185) | | (0.188) | (0.177) | |
| 5 good GCSEs incl E&M | 0.557 | 0.579 | 0.022*** | 0.603 | 0.564 | -0.039*** |
| | (0.497) | (0.494) | | (0.489) | (0.496) | |
| English score | 0.000 | -0.008 | -0.008*** | 0.120 | -0.084 | -0.204*** |
| | (1.000) | (0.990) | | (1.009) | (0.970) | |
| Math score | 0.000 | 0.087 | 0.087*** | 0.249 | -0.009 | -0.258*** |
| | (1.000) | (1.040) | | (1.033) | (1.032) | |
| N | 5,569,170 | 437,747 | | 167,148 | 270,599 | |

**Note:** Data from NPD 2007-2016. Sample of students in grades 11. Whole population is comparing pupils from minor ethnic group and the whole population. Within ethnic groups of focus is comparing pupils who ever missed for religious reasons and those who never missed. Number of observations is the max observations where different percent of missing is existing in different controls. "5 good GCSEs incl E&M" is whether achieved 5 or more GCSE and equivalents at grades A*-C/9-4(level 2) including GCSE English and Maths. GCSE test grades are transferred into scores by "U/X = 0, G = 16, F = 22, E = 28, D = 34, C = 40, B = 46, A = 52, A* = 58". Scores are standardized within year by minus mean and divided by standard deviation.

# Conclusion

This thesis has examined the interaction between school and home environments, and the impact of each on the wellbeing and outcomes of children and parents. Through three empirical studies, it contributes to a broader understanding of how educational inequalities emerge and how policy design can mitigate them.

The first chapter focuses on parental investments and explores the effects of family and private tutoring on students' school test results and standardised cognitive test performance. It finds that private tutoring improves school test rankings, particularly in mathematics, but does not enhance cognitive ability. Family tutoring, in contrast, has weak or negative effects across all outcomes. Further analysis suggests that private tutoring raises students' school exam performance by improving test-taking ability rather than cognitive ability. This aligns with existing literature showing that tutoring often benefits performance in school tests but not in cognitive assessments. Hence, educational investment may enhance students' competitive advantage in examinations without improving their underlying cognitive skills.

The second chapter examines school investments by studying the effects of China's Student Nutrition Improvement Programme (SNIP) on parents' time allocation in leisure and labour markets. It shows that the expansion of school meal provision benefits not only students but also their parents. Specifically, the availability of school meal increases urban *hukou* parents' time spent listening to music, while meal + subsidy raise rural *hukou* parents' frequency of watching television and listening to music. In terms of labour market participation, the policy significantly increases both urban and rural *hukou* parents' working hours. These findings suggest that school meal programmes designed for children can also generate welfare gains for parents and raise labour supply, especially among low-income households.

The third chapter investigates the joint effects of family and school environments on student absence and educational outcomes. It shows that a greater number of religious festival days falling on school days leads to higher absenteeism, which in turn lowers academic performance and attainment. These findings imply that predictable religious festivals could be integrated into school scheduling to reduce absences and improve learning outcomes, particularly in districts with large populations of students from minority ethnic or religious groups.

This thesis contributes to the literature in several ways. The first chapter separately identifies the impacts of private and family tutoring on cognitive skills and test-taking ability, thereby reconciling inconsistencies in previous studies. The second chapter extends the literature by examining the spillover effects of school meal policies on parental time allocation, highlighting both labour and leisure dimensions. The findings emphasise the broader socioeconomic returns of child-focused interventions and call for policy evaluations that account for both direct and indirect welfare effects. The third chapter introduces the use of festival dates as an instrumental variable in educational research, offering a new approach to understanding how religious observance influences academic outcomes. It also provides a potential policy avenue for reducing absenteeism and improving attainment among students from specific religious backgrounds.

This thesis also has several limitations. The first chapter faces challenges of reverse causality and cannot capture long-term effects. The second chapter is constrained by a small sample size, limiting the scope for further analysis. The third chapter's main limitation concerns unobserved inset days distribution, as it can weaken the link between festival timing and actual absence.

In summary, this thesis shows that education is not merely a result of individual effort or school inputs but the outcome of complex interactions between family and school environments. Across three empirical contexts, it shows how family behaviour, family characteristics, and school policies jointly shape educational success. By uncovering these dynamics, the thesis offers a more comprehensive understanding of educational inequality and provides practical insights for policies aimed at creating fairer and more effective education systems.